

THE CORRELATION BETWEEN A GENERAL CRITICAL THINKING SKILLS  
TEST AND A DISCIPLINE SPECIFIC CRITICAL THINKING TEST  
FOR ASSOCIATE DEGREE NURSING STUDENTS

Helen Reid, B.A., B.S., M.S.

Dissertation Prepared for the Degree of  
DOCTOR OF EDUCATION

UNIVERSITY OF NORTH TEXAS

May 2000

APPROVED:

Ronald W. Newsom, Major Professor and  
Program Coordinator for Higher Education

John L. Baier, Minor Professor

Jesse Jones, Committee Member

Michael Altekruise, Chair of the Department of  
Counseling, Development and Higher  
Education

M. Jean Keller, Dean of the College of Education

C. Neal Tate, Dean of the Robert B. Toulouse  
School of Graduate Studies

Reid, Helen, The correlation between a general critical thinking skills test and a discipline specific critical thinking test for associate degree nursing students. Doctor of Education (Higher Education), May 2000, 158 pp., 16 tables, 6 illustrations, references, 206 titles.

In 1997, NLNAC added critical thinking as a required outcome for accreditation of associate degree nursing (ADN) programs. Until recently general critical thinking tests were the only available standardized critical thinking assessment tools. The emphasis has shifted to discipline specific tools.

This concurrent validity study explored the correlation between two critical thinking tests, a general skills test, the California Critical Thinking Skills Test (CCTST) and a discipline specific test, the Arnett Critical Thinking Outcome Evaluation (CTOE). Both tests are based on the same definition of critical thinking. The CCTST, developed in 1990, covers discipline neutral content in multiple choice items. The CTOE, a free entry, written response test developed in 1998, assesses critical thinking in nursing situations using a partial credit model.

A convenience sample of 434 sophomore ADN students from 9 programs in Texas completed the demographic survey and critical thinking tests in 1999. The sample was 87.9% female and 74.2% Caucasian, with a mean age of 31, mean GPA of 3.13, mean 3.7 years healthcare employment experience, mean CCTST score of 15.0023 and mean CTOE of 82.69. The sample also included 22.4% current LVNs, 15.7% with prior degrees and 53.5% in the first generation of their family to go to college.

## ACKNOWLEDGEMENTS

I wish to acknowledge with special thanks and appreciation a number of individuals who provided assistance with the study. This research would not have been possible without the time and efforts of the following faculty, deans or directors from the community college associate degree nursing programs whose students participated in this study: Betty Oliver from Alvin Community College, Dr. Vivian Lilly from Collin County Community College, Joanna Barnes and Jeanie Hardin from Grayson College, Dr. Judy Walsh and Alice Myers from McLennan Community College, Maurice Robeson from North Central Texas College, Dr. Lula Pelayo and Dr. Janet Burge from San Antonio College, Dr. Sue Ochsner and Nancy Perkins from Tarrant County College, Dr. Ray Hargrove-Huttel from Trinity Valley Community College, and Dr. Edna Garza-Escobedo from the University of Texas at Brownsville/Texas Southmost College. Special thanks also go to the respondents of this study. I wish them well in their nursing careers.

I would like to gratefully acknowledge the contributions of Dr. Penelope Arnett, owner and founder of the Arnett Development Corporation, who sparked the idea for this project and provided the Arnett Critical Thinking Outcome Evaluation and evaluations free of charge to all the students participating in the study. Many thanks to Dr. Arnett's assistant, Andrea, who clarified numerous questions and worked hard to provide the data needed for the study.

I also wish to thank the members of my doctoral committee: Major Professor Dr. Ron Newsom, Program Coordinator of Higher Education; Minor Professor Dr. John Baier, Higher Education; and Dr. Jesse Jones, Executive Director of the North Texas Community and Junior College Consortium. I am grateful for their challenging, professional, and supportive encouragement as well as their attention to details.

My deepest love and gratitude go to my husband and children for their never-ending love and support. Special thanks go to my colleagues and friends for their continual support and encouragement. Above all I give thanks to God for blessing me with the gifts to complete this project and all the assistance provided along the way.

## TABLE OF CONTENTS

	Page
LIST OF TABLES .....	v
LIST OF ILLUSTRATIONS .....	vi
Chapter	
1. INTRODUCTION .....	1
The Problem	
Purposes of the Study	
Hypotheses	
Significance of the Study	
Definition of Terms	
Limitations	
Delimitations	
Assumptions	
2. LITERATURE REVIEW .....	9
Introduction to Critical Thinking and Outcomes Assessment	
Definition of Critical Thinking	
Critical Thinking as a General versus Discipline Specific Process	
Critical Thinking as an Outcome in Nursing	
Issues in Critical Thinking Assessment	
Critical Thinking Assessment Methods	
Strategies for Improving Critical Thinking in Nursing	
Standardized Critical Thinking Assessment Tools & Associated Research Studies	
The Watson Glaser Critical Thinking Appraisal	
The Cornell Critical Thinking Test	
The Ennis Weir Critical Thinking Essay Test	
The California Critical Thinking Skills Test & The California Critical Thinking Dispositions Inventory	
Other Critical Thinking Assessment Tools	
Discipline Specific Critical Thinking Assessment Tools	
Summary and Conclusions from the Literature Review	

3. RESEARCH METHODOLOGY .....	83
Research Design	
Instruments	
The California Critical Thinking Skills Test	
The Arnett Critical Thinking Outcome Evaluation Tool	
Population	
Sampling Method	
Procedure for Data Collection	
Data Analysis	
4. RESULTS .....	94
Introduction	
Descriptive Statistics of the Sample	
Critical Thinking Test Results	
Hypotheses Tested	
Additional Findings	
5. SUMMARY, CONCLUSIONS & RECOMMENDATIONS .....	119
Introduction	
Summary of Findings	
Discussion of Findings	
Conclusions	
Implications for Practice	
Recommendations for Future Study	
APPENDICES .....	131
REFERENCES .....	137

## LIST OF TABLES

Table	Page
1. Summary of Nursing Studies on Critical Thinking Using the Watson Glaser Critical Thinking Appraisal .....	45
2. Summary of Nursing Studies on Critical Thinking Using the California Critical Thinking Skills Test.....	69
3. Maximum Scores for the California Critical Thinking Skills Test .....	84
4. ANOVA Results Comparing GPA Means by School.....	101
5. Comparison of CCTST and Subtest Results .....	102
6. Comparison of CTOE Subtest Results by Number of Questions Missed.....	104
7. Comparison of CTOE Subtest Results by Raw Score .....	105
8. Regression Analysis with CCTST Scores and Independent Variables.....	113
9. Regression Analysis with CTOE Scores and Independent Variables.....	114
10. Correlation Between CCTST Total and Subtest Scores .....	115
11. Correlation Between CTOE Total and Subtest Scores .....	115
12. Summary of Stepwise Multiple Regression Analysis for CCTST.....	117
13. Summary of Stepwise Multiple Regression Analysis for CTOE.....	118
14. Summary of Results of Hypothesis One .....	120
15. Summary of Results of Hypothesis Two for CCTST Scores .....	121
16. Summary of Results of Hypothesis Two for CTOE Scores .....	122

## LIST OF ILLUSTRATIONS

Figure	Page
1. Distribution of Participants by School.....	95
2. Distribution of Ages of Students in Sample .....	96
3. Distribution of Health Care Employment Experience in the Sample .....	98
4. Distribution of GPAs of the Students in the Sample .....	100
5. Distribution of California Critical Thinking Test Scores .....	102
6. Distribution of Arnett Critical Thinking Outcome Evaluation Scores .....	104

## CHAPTER I

### INTRODUCTION

Outcomes assessment has been a criterion for accreditation of schools of nursing by the National League for Nursing (NLN) since 1991. This methodology was enacted in response to the call for accountability in higher education in the 1980s. One of the required outcomes for assessing program effectiveness in bachelors of science in nursing (BSN) programs was critical thinking. At the same time, it was an optional outcome for associate degree nursing (ADN) programs. In 1997, the National League for Nursing Accrediting Commission (NLNAC) (1997, Revised) added critical thinking as a required outcome for ADN programs. Although highly regarded as a necessary emphasis in nursing programs by faculty and employers, not much has been done to assess the status of critical thinking in associate degree nursing programs or the effectiveness of efforts to improve this ability.

The majority of the literature on critical thinking in nursing has been directed towards students in BSN programs and ways to meet the NLNAC criteria in those programs. Several cross-sectional studies compared different levels of nursing students including ADN and BSN students, but only one studied ADN students exclusively. With the addition of the critical thinking outcome requirement for accreditation, research using ADN students is imperative.

A variety of methods have been cited in the literature as being used to assess critical thinking in nursing programs, along with advantages and disadvantages of each.



Despite their shortcomings, standardized tests are popular due to their ease of administration, the length of time needed for scoring, and the ability to compare students with standardized norms. Until recently, the only available standardized tools have been general skills tests for critical thinking. The Watson Glaser Critical Thinking Appraisal (WGCTA) has been cited most often in the nursing literature but found lacking due to inconsistent results. The NLNAC (1997, Revised), in its accreditation guidelines, directs schools of nursing to define critical thinking, conduct ongoing, systematic evaluation of all program components including critical thinking, and to use the findings for development, maintenance, and revision of the program. The NLNAC does not require the use of specific tools but directs the selection of appropriate methods based on the school's evaluation plan, keeping in mind reliability, validity and specificity in the selection. Some of the evaluation methods mentioned are portfolios, performance observation, graduate surveys, simulated clinical situations, and scores on standardized or locally developed tests. The only tests specifically listed in the accreditation guidelines are the California Critical Thinking Skills Test (CCTST) and the California Critical Thinking Dispositions Inventory (CCTDI). Created in 1990 and 1992, these tests have begun to be used for a variety of assessment studies. Research in the nursing literature using the CCTST also provides inconsistent results.

Besides the general psychometric issues such as reliability, validity, and availability of appropriate norms, nursing faculty must consider whether to use a general critical thinking skills test or a discipline specific tool. Until recently, there have not been any discipline specific critical thinking tests for nursing. The Arnett Critical Thinking Outcome Evaluation (CTOE) tool, a discipline-specific, constructed response test for

nursing, became available in 1998. The CTOE results include a total score and subscale scores for each of the five cognitive skills of analysis, interpretation, evaluation, explanation, and inference. The CCTST is a general critical thinking skills test. Scores for the CCTST include a total score, and subscale scores for analysis, evaluation, inference, deduction and induction.

Both the Arnett CTOE tool and the CCTST are based on the consensus definition of critical thinking from the Delphi Report. The Delphi Report is the result of the two year work of a national panel of experts across the United States and Canada using the Delphi methodology to achieve consensus about what to expect in critical thinking of college freshmen and sophomores (P. A. Facione, 1991).

Studies have reported correlations among several different critical thinking tools. The relationship between the CCTST and the CCTDI ( $r = .667$ ) and the relationship between the CCTST and the WGCTA ( $r = .405, .544$ ) have been studied (Facione & Facione, 1998; Facione, Facione, & Sanchez, 1994). The lower correlation between the WGCTA and the CCTST is explained by the fact that they are based on different definitions of critical thinking. No studies could be found examining the relationship between a general critical thinking skills test and a discipline specific skills test. Because of the controversy over the use of general critical thinking skills tools versus discipline specific critical thinking tests, and the need for associate degree nursing programs to choose methods for evaluating critical thinking, more information was needed about both methods.

## The Problem

What level of concurrent validity exists between the California Critical Thinking Skills Test, a general critical thinking skills test, and the Arnett Critical Thinking Outcome Evaluation, a discipline specific critical thinking test?

## Purposes of the Study

This study explored the correlation between California Critical Thinking Skills Test scores and Arnett Critical Thinking Outcome Evaluation scores of associate degree nursing students to determine if there was evidence of concurrent validity between the two instruments. These tests, a general critical thinking skills test and a discipline specific critical thinking test, are based on the same definition of critical thinking. A secondary purpose of this study was to determine if any of several variables may be associated with critical thinking scores for associate degree nursing students.

## Hypotheses

The hypotheses for this study were as follows:

H 1: There is a correlation between scores of associate degree nursing students on a general test of critical thinking skills, the CCTST, and a discipline specific critical thinking test, the Arnett CTOE tool, in the following ways:

- a. There is a positive correlation between the total scores of associate degree nursing students on the CCTST and the total scores of the Arnett CTOE tool.
- b. There is a positive correlation between the analysis subscale scores of associate degree nursing students on the CCTST and the combined analysis and interpretation subscale scores of the Arnett CTOE tool.

- c. There is a positive correlation between the evaluation subscale scores of associate degree nursing students on the CCTST and the combined evaluation and explanation subscale scores on the Arnett CTOE tool.
- d. There is a positive correlation between scores of associate degree nursing students on the inference subscale scores on the CCTST and the inference subscale scores on the Arnett CTOE tool.

H 2: Intervening variables are associated with critical thinking in associate degree nursing students in the following ways:

- a. Age is positively correlated with critical thinking scores on a general critical thinking skills test and a discipline specific critical thinking test for associate degree nursing students.
- b. Completion of a previous degree is positively correlated with critical thinking scores on a general critical thinking skills test and a discipline specific critical thinking test for associate degree nursing students.
- c. GPA is positively correlated with critical thinking scores on a general critical thinking skills test and a discipline specific critical thinking test for associate degree nursing students.
- d. First generation in college status is negatively correlated with critical thinking scores on a general critical thinking skills test and a discipline specific critical thinking test for associate degree nursing students.
- e. The number of years experience working in a health care setting is positively correlated with critical thinking scores on a discipline specific critical thinking test

- but is not correlated with critical thinking scores on a general critical thinking skills test for associate degree nursing students.
- f. There is no correlation between a student's status as an LVN and critical thinking scores on a general critical thinking skills test and a discipline specific critical thinking test for associate degree nursing students.
  - g. There is no correlation between gender and critical thinking scores on a general critical thinking skills test and a discipline specific critical thinking test for associate degree nursing students.
  - h. There is no correlation between ethnicity and critical thinking scores on a general critical thinking skills test and a discipline specific critical thinking test for associate degree nursing students.
  - i. There is a statistically significant correlation between the variable set including age, completion of a previous degree, and GPA and critical thinking scores on a general critical thinking skills test and a discipline specific critical thinking test for associate degree nursing students.

### Significance of the Study

The significance of this study is three-fold. First, information is given regarding the psychometrics of the two critical thinking assessment tools. Detailed information about such tools is helpful to those making the decision at each associate degree nursing school about what instrument to use to measure critical thinking to meet accreditation criteria. With the revised criteria taking effect in 1997, the timing is of great significance to ADN programs. Secondly, the statistics regarding the relationship between a general

critical thinking skills test and a discipline specific skills test is useful as part of the information faculty use to decide whether to use a general or discipline specific skills test. Thirdly, this study adds to the body of knowledge regarding critical thinking in associate degree nursing students. With the majority of the information about critical thinking in nursing being available at the baccalaureate degree level, and the inconsistency of the results obtained to date, it is all the more important to gather additional data.

#### Definition of Terms

The following definitions apply to this study:

1. General critical thinking skills test – California Critical Thinking Skills Test (CCTST)
2. Discipline-specific critical thinking skills test – Arnett Critical Thinking Outcome Evaluation (CTOE)
3. Critical thinking- “purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based” (Facione, 1990, p. 3)
4. Associate Degree Nursing program – a two-year program of study in nursing usually at a community or junior college, leading to an associate degree. Graduates of ADN programs take the NCLEX-RN, the licensing exam, to become registered nurses.

5. Previous degree – attainment of an associate, bachelors or masters degree from an institution of higher education prior to enrolling in the current associate degree nursing program

6. First generation at college status – student's generation is the first one to have a member go to and graduate from college of any kind

7. LVN status – whether a student is or is not a licensed vocational nurse prior to enrolling in the current associate degree nursing program

#### Limitations

The following limitations applied to this study:

The convenience sampling procedure decreases the generalizability of the findings.

#### Delimitations

The following delimitations applied to this study:

A delimitation of this study was that it only tested associate degree nursing students in NLNAC accredited programs in the state of Texas. Another was that this study only used the CCTST and the Arnett CTOE.

#### Assumptions

The following assumptions were made for this study:

1. Critical thinking can be defined and measured.
2. Students will be equally motivated to perform on both tests.

## CHAPTER II

### LITERATURE REVIEW

#### Introduction to Critical Thinking and Outcomes Assessment

Within the last two decades, critical thinking has become a national imperative for higher education. This outcome received a major surge into national emphasis following three major reports on American education in the middle 1980s, “A Nation at Risk” from the National Commission on Excellence in Education in 1983, “Involvement in Learning” from the National Institute of Education in 1984, and “Higher Education and the American Resurgence” in 1985 (Brookfield, 1987). The 1980s were characterized as the assessment decade (Yarbrough, 1992). The nation's governors listed critical thinking as a core skill in their National Educational Goals (Tucker, 1996b). In 1983, the U.S. Department of Education called for a great increase in the percentage of college graduates with higher critical thinking ability in Goal 5.5 of the national Educational Goals 2000 (N. C. Facione, 1995; Facione & Facione, 1994; Terenzini, Springer, Pascarella, & Nora, 1995). A reversal of the emphasis on lower order thinking skills was needed (Paul & Nosich, 1992). In addition to the government, corporate America began advocating improved ways of teaching and assessing thinking skills to enhance business productivity. In “America’s Choice: Higher Skill or Low Wages” from the 1990 SCANS project, critical thinking was included as one of the core skills (Tucker, 1996b).

At the same time, a pervasive emphasis on accountability developed nationally. One way that institutions of higher education responded to accountability demands was



by shifting to an outcomes-based evaluation or outcomes assessment (Halpern, 1987; Miller, 1992). Outcomes assessment involves choosing and clearly identifying a construct, specifying behavior that clearly elicits that construct, choosing methods to assess the behavior, measuring that behavior, and interpreting the behavior that was demonstrated. An appropriate gauge must be available to interpret the results (Davis, 1991; Whitney, 1993).

There are several outcomes assessment programs on critical thinking taking place on college, city, state, and national levels. The majority of these are focused on primary and secondary education such as the Pittsburgh Public Schools Monitoring Achievement in Pittsburgh Critical Thinking Project (Moss & Koziol, 1991), the New Jersey statewide assessment of students' thinking skills (Morante & Ulesky, 1984), and the Connecticut Reasoning Mastery Test (Sternberg, 1985). Assessment of critical thinking skills was mandatory for college graduates in eight states in the early 1990s (N. C. Facione, 1995). California regulations mandated that all Associate of Arts degree courses require critical thinking ability (Coffman, 1988). DeVry Education is adding critical thinking assessment in its occupational programs (Tucker, 1996b). The Office of Statistics commissioned proposals for a national assessment plan to transform national objectives into a specific plan (Halonen, 1995). Paul and Nosich (1992) have been working on a Model for National Assessment of Higher Order Thinking that will provide national standards in higher order thinking as a motivation for and a guide to instruction. To make the program affordable, testing skills, abilities and dispositions would be done on a representative sample of students at various points from primary through post-secondary education with a combination of multiple choice items, multiple rating items, short essay, and

constructed response items (Paul & Nosich, 1992). Similar assessments are taking place in other countries. In a formal evaluation of a broad-based, cross-disciplinary critical thinking program in Venezuela, students outperformed controls on a standardized test of thinking skills (Halpern, 1993).

### Definition of Critical Thinking

Critical thinking is an abstract, complex concept that is difficult to define. It is important that the definition of critical thinking can be operationalized, leading to useful measures and predictors of behavior. It should fit with modern theories of intellect (Tucker, 1996a). There are multiple definitions of critical thinking in the literature. Many have areas of overlap, and some have a narrow focus and others a broad view. Definitions tend to have philosophical, psychological or educational orientations (Sternberg, 1986). Paul (1990) suggests that too much emphasis should not be put on any one definition. Many definitions are needed to cover the different aspects of critical thinking.

In 1941, Glaser described three components of critical thinking in his definition, which included an attitude or disposition to evaluate problems in a thoughtful analytical way, knowledge of the methods of inquiry and reasoning, and skill in applying the measures (Cross & Steadman, 1996). In many of the definitions that have followed, emphasis has been on two components that combine Glaser's three aspects. Examples of these two components from various studies are cognitive skills and affective dispositions (Baker, 1996; Norris & Ennis, 1989; Paul, 1990; Sternberg, 1986; Videbeck, 1997a), skills to process and generate information and habits of using these skills to guide behavior (Paul & Elder, n.d. 1998, March 7; Scriven & Paul, 1996), reasoned judgment

and reflective thinking (Jones & Brown, 1993; Norris & Ennis, 1989), a cognitive process grounded in reflection (Baker, 1996), reflective judgment (King & Kitchener, 1994), and a multidimensional, metacognitive activity (Beeken, Dale, Enos, & Yarbrough, 1997; Kintgen-Andrews, 1991). The disposition to engage in critical thinking or critical thinking attitude is seen by many to be as important as the cognitive skills (Halpern 1989). Others talk about critical thinkers having a “critical spirit” (P. A. Facione, 1996; Kintgen-Andrews, 1991; Norris, 1985; Norris & Ennis, 1989). Several of the authors have created lists of cognitive and affective skills (Ennis, 1985; P. A. Facione, 1991; Norris & Ennis, 1989; Paul, 1990). Some of the skills included are identifying and challenging assumptions (Beeken, Dale, Enos, & Yarbrough, 1997; Brookfield, 1987; Ennis, 1993; P. A. Facione, 1991; Terenzini, Springer, Pascarella, & Nora, 1995), exploring and imagining alternatives (Brookfield, 1987; Christen, Angermeyer, Davison, & Anderson, n.d. 1998, March 7), developing reflective skepticism (Brookfield, 1987), active inquiry (Brookfield, 1987), evaluation of interpretations (N. C. Facione, 1995), and making correct inferences (Terenzini, Springer, Pascarella, & Nora, 1995).

There are common threads in many of the definitions of critical thinking. One of Paul’s definitions of critical thinking is “the art of thinking about your thinking while you’re thinking so as to make your thinking more clear, precise, accurate, relevant, consistent and fair” (Paul, 1990, p. 32). This focus on metacognition and self-regulation is found in many definitions. Metacognition was also found in Dewey’s (1933) reflective thinking. Ennis defines critical thinking as “reasonably reflective thinking focused on deciding what to believe or do” (Ennis, 1993, p. 180). Facione and Facione (1996a) characterize critical thinking as a nonlinear, recursive process resulting in judgments

about what to believe or do in a given context. The nonlinearity and contextual nature of critical thinking are other recurring themes found in many definitions (Alfaro-LeFevre, 1995; Facione & Facione, 1996a; Jacobs, Ott, Sullivan, Ulrich, & Short, 1997; Paul, 1990). Another thread gradually entering definitions of critical thinking is its collaborative nature (Barnett, 1997). Paul (1990) also mentions the need to engage in dialogical and dialectical reasoning. Many issues require dialectical thought that crosses between disciplines. Critical thinking has also been called higher order or multilogical thinking. An essential element of critical thinking is seeing things from another's point of view, or dialogical thinking (Sternberg, 1986).

In 1987, the American Philosophy Association created a multidisciplinary committee of forty-six published critical thinking theorists from philosophy, education, and psychology, including the notable Robert Ennis, Stephen Norris, and Richard Paul (P. A. Facione, 1996). This committee conducted a Delphi research project leading to the development of a consensus definition and conceptualization of critical thinking and its core cognitive skills and affective dispositions (Colucciello, 1997; N. C. Facione, 1995). This definition, known as the Delphi definition, has been called the “conceptual architecture” used to achieve national consensus among hundreds of educators, employers and policy makers (Facione & Facione, 1996a). The final definition of the project was that critical thinking is:

...purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation and inference, as well as explanation of the evidential, conceptual, methodological, criteriological or contextual considerations upon which that judgment is based. (Facione, 1990, p. 3)

This definition lists the consensus critical thinking cognitive skills of interpretation, analysis, evaluation, inference, explanation and self-regulation (P. A. Facione, 1996; Facione & Facione, 1996a). Defined criteria for each cognitive skill help direct teaching and evaluation of critical thinking skills (P. A. Facione, 1996). Dispositions included in the final report were inquisitiveness, open-mindedness, analyticity, critical thinking self-confidence, systematicity, cognitive maturity, and truth seeking (N. C. Facione, 1995). A group of faculty at eighty colleges, government workers and employers endorsed this definition after a replication of the Delphi study was commissioned by the federal government at Penn State University's National Center on Post-Secondary Teaching, Learning, and Assessment (N. C. Facione, 1995).

#### Critical Thinking as a General versus Discipline Specific Process

With all the areas of consensus, there is still a major controversy regarding whether critical thinking should be defined as subject matter specific or a general process. There is an ongoing debate regarding the ability to think apart from specific subject content. The majority of the experts on critical thinking support emphasis on critical thinking as a general process, free of the specific contexts in which it is manifested. Nummedal (1991) believes that most conceptualizations of critical thinking are too narrow and discipline specific. Focus should instead be on critical thinking skills relevant to everyday issues. Alternatively, McPeck (1981) believes that critical thinking is the "propensity and skill to engage in an activity with reflective skepticism" (p. 8) and argues that it should be discipline specific and taught within the field. He believes that one does not think critically without thinking about something. Critical thinking has been

described as the cognitive engine driving knowledge development and professional judgment in a wide variety of professional practice fields (Facione & Facione, 1996a). Different disciplines foster unique aspects of critical thinking (Cross & Steadman, 1996) and require different kinds of reasoning skills or emphasize certain generic skills differently (Kintgen-Andrews, 1991). Tucker (1996a) does not see much equivalence between critical thinking in diverse disciplines. Many issues related to critical thinking cannot be solved until there is more evidence about what it means in various professional contexts and what aspects can be changed with education. Colucciello (1997) has called for the development of criteria for evidence of critical thinking in each discipline.

#### Critical Thinking as an Outcome in Nursing

A parallel focus on critical thinking developed in nursing. In 1934, Beck (cited in Hanson, 1991) described critical thinking as a scientific attitude that nurses needed to develop. In 1937, the National League for Nursing (NLN) noted that the courses comprising liberal education in a bachelor of science in nursing (BSN) curriculum provided the foundation for development of critical thinking (Hartley & Aukamp, 1994). Critical thinking was added as a topic in the Cumulative Index for Nursing and Allied Health Literature (CINAHL) in 1989 (Hartley & Aukamp, 1994).

The NLN Accreditation Outcomes Project reported that critical thinking was on many of the schools of nursing top ten lists of outcomes that should be assessed, but not on the top ten lists of those actually being measured (Hickman, 1993). Few programs actually measured critical thinking, and when they did, few measured the validity and reliability of instruments used (Rane-Szostak & Robertson, 1996). In keeping with

national trends, the NLN, which accredits nursing programs, added outcomes assessment as an accreditation criterion in 1991. In this criterion, the standard for educational effectiveness requires that the program have an identified plan for evaluation and feedback to strengthen its educational outcomes, including critical thinking. The accreditation process requires that each program define critical thinking as one of several outcomes, give rationale for the ways it is measured, and use the evaluation results for development, maintenance, and revision of the nursing program (National League for Nursing Accrediting Commission, 1997, Revised).

Although the NLN saw the associate degree nursing (ADN) graduate's practice as characterized by critical thinking, critical thinking was originally included as a required outcome only for BSN programs. Both ADN and BSN programs are required to include strategies to improve critical thinking skills of their students (Alfaro-LeFevre, 1995). The emphasis on critical thinking was thought to be appropriate for BSN graduates because of their greater scope of practice and responsibilities (Miller & Malcolm, 1990). Some argue that critical thinking skills are effectively taught only at the BSN education level, but both ADN and BSN graduates practice under the same license as registered nurses, and are held to the same level of responsibility and accountability (Brooks & Shepherd, 1992). Nurses need critical thinking skills to be safe, competent practitioners (Miller & Malcolm, 1990). Students must demonstrate skills in reasoning, analysis, and research or decision-making relevant to the nursing discipline to meet the NLN outcome criteria (Alfaro-LeFevre, 1995; Howenstein, Bilodeau, Brogna, & Good, 1996; Vaughan-Wrobel, O'Sullivan, & Smith, 1997). Because nursing practice requires that nurses have skills to interpret, analyze, and formulate action based on more and more information, the

National League for Nursing Accrediting Commission (NLNAC) (1997, Revised) added critical thinking as an outcome for ADN programs as well as BSN programs in the criteria for accreditation when they were revised in 1997.

The inclusion of critical thinking outcomes and assessment measures into BSN programs has been a complex and difficult task (Dexter, Applegate, Backer, Claytor, Keffer, Norton, & Ross, 1997). Much has been written about the need for critical thinking in nursing, but actual research is not plentiful (Miller, 1992). There is great concern about the use of critical thinking as an outcome measure when adequate means of testing it have not yet been developed (Hickman, 1993). Nursing faculty also have difficulty operationally defining critical thinking within the curriculum (Colucciello, 1997).

A descriptive study by Jones and Brown (1991) of the conceptualization of critical thinking by 470 deans and directors of BSN programs found divergence exists in defining and understanding critical thinking. Critical thinking was conceptualized as a variation of the scientific method, a logical, rule-driven decision making process similar to the nursing process, where judgment and skepticism are discouraged. It was operationalized as a rational linear process. The survey results showed that eighty percent of the responding deans and directors thought critical thinking was similar to the nursing process because analysis and evaluation, two components of the nursing process, are important components of critical thinking (Perciful & Nester, 1996). Critical thinking has traditionally been equated with the nursing process, which many believe is the framework within which nurses apply critical thinking skills (Miller & Malcolm, 1990). When using the nursing process, nurses must use inductive and deductive reasoning, make hypotheses, plan, implement, and evaluate patient care (Howenstein, Bilodeau, Brogna,



& Good, 1996). Many of the cognitive problem solving skills needed to implement the nursing process are those defined as critical thinking abilities. Perciful and Nester (1996) report that the literature suggests the nursing process hinders critical thinking. Decision making in nursing is composed of contextually defined value judgments (Glen, 1995). Faculty are increasingly recognizing the limitations of the nursing process for teaching and assessing critical thinking and clinical judgment (Gendrop & Eisenhauer, 1996, Tanner, 1993). Tanner (1997) and Woods (1993) note that faculty in nursing programs often emphasize reason and logic in critical thinking with little attention to the affective component. The nursing process may be necessary but it is insufficient (Gendrop & Eisenhauer, 1996; Glen, 1995; Jones & Brown, 1993). Less than 50% of the surveyed deans and directors supported the role of context in critical thinking, which goes against current trends in the literature (Gendrop & Eisenhauer, 1996). In nursing practice, problems are rarely solved in a linear manner and critical thinking involves contextually defined value judgments (Jones & Brown, 1991). Recent models of critical thinking keep problem solving as a focus and integrate issues of development, context and reflection (Gendrop & Eisenhauer, 1996).

O'Sullivan, Blevins-Stephens, Smith, and Vaughan-Wrobel (1997) conducted a similar survey to that of Jones and Brown, and added master of science in nursing (MSN) programs. A random sample of critical thinking definitions used in nursing programs showed that a high percentage (37%) conceptualized critical thinking as scientific problem solving, a linear process. Close to 13% used a definition from the National Council of Excellence in Critical Thinking, 9% used the Watson-Glaser definition, 6% used definitions from critical thinking experts, and 13% used a nursing specific

definition. In 1994, three years after the NLN added critical thinking as a required outcome for BSN programs, only 69% of the BSN and 22.5% of the MSN programs had developed definitions. Identified difficulties in operationalizing critical thinking were definition, measurement, faculty attitudes and skills, synthesis process, time, cost, and logistics.

Videbeck (1997b) conducted a descriptive study using NLN self-study reports from 55 schools of nursing to describe current practice. For the definition of critical thinking, 43 programs included cognitive and affective aspects, and 12 had only cognitive components. Fifty schools used an adaptation or adoption of a particular author's definition of critical thinking, with 28 being nursing oriented and 22 non-nursing oriented. Of the definitions used, 11 chose Bandman and Bandman, 10 chose Miller and Malcolm, 8 chose Watson and Glaser, 8 chose Paul, and 5 chose Ennis.

Gordon (1997) examined the relationship between conceptualization of critical thinking of nurse educators in the midwest and critical thinking experts in an exploratory study. A significant difference was found among their perceptions. Nurse educators were more likely to regard decision-making, reasoning, and problem solving as critical thinking activities than the critical thinking experts were.

An example of implementation of the required critical thinking outcomes assessment is offered by the Indiana University School of Nursing which has four levels of nursing programs from ADN to doctorate. The faculty believe that rational thinking involves the same elements at every educational level with differences in the level of reasoning expected at each level. An across program critical thinking outcome based on the Delphi project critical thinking definition was adopted. Competency statements for

each of the four levels were developed and recommendations regarding measurement were made. Since they could not find a critical thinking measurement instrument with sufficient reliability and validity in the nursing context they decided to use the California Critical Thinking Skills Test (CCTST) and California Critical Thinking Dispositions Inventory (CCTDI). Students will be measured at different points to see if significant changes occur over time and to see how well the scores correlate with clinical performance measures, course grades, GPA, and NCLEX results. If these tests are found sufficiently valid and reliable in the nursing context they will be adopted for evaluation purposes. Their assessment plan includes evaluation of the students within nursing contexts and with multiple measures including direct observation and course specific written tests that require critical thinking. No outcomes were reported (Dexter, Applegate, Backer, Claytor, Keffer, Norton, & Ross, 1997).

### Issues in Critical Thinking Assessment

There are several issues that must be addressed before developing a critical thinking assessment program. Assessors must be aware of the strengths and weakness of the assessment approaches they choose. Much attention has been given to the definition of critical thinking, and designing and implementing strategies to improve it, but little has been done to see if these efforts are truly effective. For students to be assessed on the critical thinking outcome, evidence of critical thinking must be externalized (N. C. Facione, 1995). Pascarella and Terenzini (cited in Cross & Steadman, 1996) note that the reliable and valid measurement of higher order thinking is a difficult and highly complex psychometric task.

Developing a distinct definition of critical thinking is the first issue to be tackled in outcomes assessment (Rane-Szostak & Robertson, 1996). The definition chosen by the individual program needs to be based on its goals and objectives. This definition will direct the assessment process (Carpenter & Doig, 1998). Any critical thinking assessment needs to be based on an operational definition of critical thinking (Halpern, 1993). The program should carefully choose a critical thinking assessment tool that was developed based on a definition and conceptualization of critical thinking that is congruent with that of the school of nursing (Cromwell, 1992; Facione & Facione, 1994; Payne, Vowell, & Black, 1991; Rane-Szostak & Robertson, 1996). The definitions on which the program and instrument are based should also match classroom activities (Christen, Angermeyer, Davison, & Anderson, n.d. 1998, March 7). Several of the research studies in the nursing literature suggest that findings were not significant because the tool was based on a different definition of critical thinking than that which they were trying to measure. Instruments are based on slightly different theoretical constructs, with different scopes that limit concurrent validity expected between them (Facione & Facione, 1994).

Rane-Szotstak and Robertson (1996) stress the need for a clear nursing focused definition of critical thinking. In nursing, clinical judgment, clinical reasoning and critical thinking are often considered synonymous (Alfaro-LeFevre, 1995). Alfaro-LeFevre (1995) defines clinical judgment as critical thinking in the clinical area. Zbilut (1995) says that critical thinking is a process and clinical judgment is content. Most definitions are global and include use of the nursing process, problem solving and decision-making (Videbeck, 1997b). Nursing needs a broader definition that includes problem solving, reasoning in considering multiple viewpoints, clinical judgment, creativity, and attitudes

of inquiry (Loving, 1993; Rane-Szostak & Robertson, 1996; Tanner, 1996). Critical thinking in nursing emphasizes context and reflection and is nonlinear as nurses must work within the patient's frame of reference (James & Clarke, 1994). When critical thinking is seen as only problem solving, the philosophy in use is a curriculum-as-product orientation. The new model for critical thinking is a curriculum-as-praxis orientation. Praxis is a form of action and reflection. Critical thinking is a process in which knowledge and action are dialectically related through mediation of critical reflection (Ford & Profetto-McGrath, 1994). Colucciello (1997) believes the Delphi definition allows nursing to move beyond the narrow linear models currently used.

Next, attention must be given to the purpose of the assessment. Purposes may include program evaluation and improvement, research, individual student achievement with diagnosis of strengths and weaknesses, collection of data for accountability, and assisting with decisions regarding an individual student's admission to certain programs (Ennis, 1993). High stakes purposes require a more accurate measurement of critical thinking abilities in order to make valid decisions. Most of the current tools should not be used for high stakes purposes (Ennis, 1993). If decisions are to be made based on critical thinking assessment, information must be of high quality, and include control of competing variables or rule out competing causes. The instruments chosen for outcomes assessment must be suitable for their intended purpose. No test is suitable for all purposes (Norris & Ennis, 1989).

It is difficult to develop a reliable and valid tool that assesses actual gains in critical thinking. Critical thinking, as measured by some of the comprehensive tests, usually changes over a long period of time, and is influenced by many factors. Some

tools actually measure aptitude that does not change much over time (Rane-Szostak & Robertson, 1996). Many of the critical thinking assessment studies look for an expected change in a relatively short time frame, which may be unrealistic (Cross & Steadman, 1996; Ennis, 1993; Halpern, 1993; Kintgen-Andrews, 1991; Schank, 1990; Terenzini, Springer, Pascarella, & Nora, 1995). A tool is needed that is sensitive enough to detect small changes in critical thinking skills and dispositions following instruction designed to improve critical thinking. Multiple comparison groups are needed to separate out maturational gains and general improvement in thinking which may result from the college experience itself (Halpern, 1993).

Another issue in critical thinking assessment is deciding between the use of norm-referenced versus criterion-referenced tools. The majority of the currently available tools are norm-referenced. These tools show the overall distribution of scores and provide comparison with other groups but there is no guarantee that the students have met a certain level of competency. Rane-Szostak and Roberston (1996) categorize norm-referenced tests as weak measures of critical thinking. Criterion-referenced measures are performance-based and more useful for measuring critical thinking ability.

An additional feature of critical thinking assessment tools is whether they are aspect specific, measuring a single subject or task, or comprehensive, measuring several different areas of critical thinking. Arter (1987) characterizes this dichotomy as atomistic versus holistic. The tools need to be carefully investigated. Critical thinking tests can have great differences between them even though they each are called critical thinking tests (Arter, 1987). With comprehensive tools, greater diagnostic ability is lost compared to the aspect specific tools (Norris & Ennis, 1989).

When selecting critical thinking assessment tools, faculty must decide if general knowledge or subject specific tests will be better suited for the selected purposes. General knowledge tests do not require specialized knowledge of a particular discipline, whereas subject specific tests measure critical thinking in certain subjects or disciplines. The Committee of the National Academy of Education has recommended the development of subject specific higher order thinking tests (Ennis, 1993). No subject specific critical thinking tests were found in articles published in 1989 (Norris & Ennis, 1989), 1993 (Ennis, 1993) or 1997 (Videbeck, 1997b). Several discipline specific tests have been developed since then, with at least two in nursing (Arnett Development Corporation, n.d. 1998, March 7; [InterEd@InterEd.com](mailto:InterEd@InterEd.com), n.d. 1998b, March 9). There is a concern with subject specific tests, that interest in, and familiarity with, the subject matter influences one's critical thinking scores (Kintgen-Andrews, 1991). Norris and Ennis (1989) believe that critical thinking should be evaluated in both discipline specific contexts and everyday life contexts. The trend has been toward subject specific, proficiency-based alternative assessments (Davis, 1991).

Tucker (1996a) notes that current critical thinking assessment tools measure skills that are necessary for critical thinking but may not be sufficient for it. The current generalist approach is based on the assumption that having foundational critical thinking skills will ensure that they can be contextualized and resolved. This may be a faulty assumption. Tucker (1996a) believes that a correct resolution of a contextualized case-oriented discipline specific problem makes the reverse assumption that the foundational skills are present and is a better approach to critical thinking assessment. A possible

negative result of such an assessment is that the scores may be diagnostically weak and will not give useful feedback.

Another issue regarding use of a critical thinking assessment tool is its usability, referring to such variables as cost, faculty and personnel resources, and availability of equipment for administering, scoring, and/or data analysis (Arter, 1987; Rane-Szostak & Robertson, 1996). Logistical aspects include availability of the test, length of testing time, ease of administration, availability of norms and guidelines for interpretation, and acceptance by stakeholders (Arter, 1987). Arter (1987) provides a checklist for selecting a critical thinking test, including examining areas of usefulness and technical analysis. It is important that the information obtained from the tool matches what is desired. Tucker (1996a) notes that with a sizable budget, eight to ten hours of assessment time and highly motivated students, critical thinking can be assessed well, but this is not very realistic. The most useful tests are short, interesting, and inexpensive, with efficient questioning methods.

Another concern regarding critical thinking tests is that of fairness. The tool should not unfairly disadvantage or advantage any groups based on reading ability, knowledge of content, gender, age, ethnicity, socioeconomic status, or cultural assumptions (Ennis, 1993, P. A. Facione, 1996). A problem noted with several standardized tests is confounding background beliefs with critical thinking ability. Using verbal reports of thinking to develop tests can lessen this (Norris, 1988).

An additional issue that has been raised in several articles is that of student motivation to do well on the assessment tools. Payne, Vowell, and Black (1991) note that student motivation is inconsistent. Tucker (1996a) notes that students are over-assessed



and their motivation to do well can no longer be taken for granted. Student motivation for tests such as the GRE is intrinsic, but may be low for tests conducted only for program evaluation or institutional assessment (Banta, Lund, Black, & Oblander, 1996). Several suggestions have been made to improve this motivation such as giving rewards, and making it a requirement for the class or graduation (Payne, Vowell, & Black, 1991).

Other issues cited in the literature include the lack of control groups in many research studies, the dropout problem, and students becoming test weary or test wise by graduation (Ennis, 1993; Halpern, 1993). Studies need to control for other possible variables leading to differences in critical thinking scores (Ennis, 1993). There is an identified need for evidence of the long-term impact of critical thinking instruction and for evidence of effects of specific teaching strategies if diagnosis and remediation of weaknesses is a goal of critical thinking instruction (Norris, 1985).

The two main indicators of quality of a critical thinking assessment tool that must be considered are its reliability, or consistency, and its validity, or trustworthiness. Commercial critical thinking tests have the advantage of having reliability and validity reported (Rane-Szostak & Robertson, 1996). Reliability estimates and validity coefficients may be low because of the complexity and the many dimensions of critical thinking (Frisby, 1991).

For reliability, measures of internal consistency, such as split-half reliabilities, Kuder-Richardson 20 and Coefficient Alpha, test-retest reliabilities, and parallel form reliabilities may be reported. For critical thinking assessment tools, reliability is the extent to which individual differences in scores are attributable to true differences in critical thinking abilities and not due to chance error. The reliability or consistency

between two sets of scores is expressed in terms of correlation coefficients (Anastasi, 1982). Anastasi (1982) notes that correlation coefficients in general should be in the .80s or .90s but for group use, Norris and Ennis (1989) note that reliability should be above .7. Most overall critical thinking assessment tool reliabilities are reasonable but lower than other standardized tests, such as intelligence tests, at about .65 to .75. Norris & Ennis (1989) recommend that this be considered adequate for critical thinking tools as there is no reason to believe that all items will correlate highly with one another. The K-R 20 is the consistency of responses to all items in a test. The more homogeneous the domain being tested, the higher the interitem consistency (Anastasi, 1982). Critical thinking is a highly heterogeneous criterion. Thorndike (1967) noted that if a test has heterogeneity in content, the K-R 20 would provide an underestimate of the correlation coefficient. For open-ended critical thinking assessment tools, interrater reliability is also of concern and must be addressed along with other forms of reliability. Reliability coefficients should be reported in the test manual accompanying the test, along with a description of the type of group on which it was determined (Anastasi, 1982).

Most critical thinking subtest reliabilities tend to be low (Arter, 1987). Tucker (1996b) believes that the greatest limitation of current critical thinking tests is that they do not possess an underlying empirical structure needed to support the instrument's a priori definition of subtests and the constructs they define. When factor analysis is done on scores from instruments claiming to assess five or more critical thinking skills, evidence is not produced that these skills sets exist or are measured by the instrument (1996a). Empirically derived subtests are needed to establish discriminant validity. The APA Delphi Report states that the critical thinking cognitive skills act in interdependent

and interconnected ways. Facione and Facione (1994) note that a factor analysis performed to differentiate the different skills into factors will likely fail because the skills have to be used interactively to answer the items. They suggest that the subscale scores should not be used for more than gross indicators of possible critical thinking strengths and weaknesses.

There are several types of validity reported for critical thinking assessment instruments. It is important that there is evidence that the tool measures what it claims and not something else (Arter, 1987; Norris & Ennis, 1989). Validity is the central concept of test theory (Loevinger, 1967). The three basic types of validity relevant to critical thinking assessment tools are content, criterion, and construct. Content validity is common for critical thinking tests. Content validity depends on relevance of responses to the behavior being measured rather than on the relevance of the item content. The test manual should describe the procedures followed in test construction to assure that the items are an adequate sample of the universe in which the researcher is interested, if subject matter experts contributed to test construction or validation, and what their qualifications were (Cronbach & Meehl, 1967; Anastasi, 1982). A review of the literature, determining the test's relationship to a theoretical model, or a review by knowledgeable critical thinking experts helps determine content validity (Arter, 1987; P. A. Facione, 1996; Norris & Ennis, 1989).

Criterion validity includes predictive and concurrent validity. Performance is measured against a criterion, a direct and independent measure of that which the test is designed to predict. Tests may be validated against many criteria. Correlation between a new test and a previously available test is frequently cited as evidence of concurrent

validity (Anastasi, 1982). Existing critical thinking tests highly correlate with verbal abilities, making it harder to distinguish critical thinking from verbal intelligence (Sternberg, 1986). Tucker (1996b) believes that the link between any of the critical thinking models and current measures of critical thinking is very weak. The current critical thinking tests provide no more than a weak to moderate substitution for IQ and/or achievement tests. He advises considering alternatives before making important curricular or programmatic decisions based on scores from critical thinking tests.

The most crucial issue for critical thinking tools is construct validity. It is more difficult to measure a process than a product. Many claims of validity rest on judgments of field experts rather than objective empirical research (Frisby, 1991). When there is no adequate criterion or direct measure to allow for an operational definition of a behavior, such as critical thinking, indirect validation measures must be used to determine what construct accounts for test variance. Construct validity is the extent to which a test measures a theoretical construct or trait (Anastasi, 1982). The underlying trait is more important than the test items or the test scores (Cronbach & Meehl, 1967). Construct validity is established by convergence of several pieces of evidence. There are several ways to experimentally measure construct validity. One is to test for group differences in groups that are expected to differ. Another is factor analysis. Studies that measure change over time, especially retesting after intervention, add to the validity. Many types of evidence are relevant to construct validity, including content validity, criterion validity, correlation of subtests with the whole, and correlation with expectations (Cronbach & Meehl, 1967). If a test correlates with variables that it should and does not correlate with variables with which it should differ, convergent and divergent validation

is shown. Construct validity cannot generally be expressed in a correlation coefficient. It requires a gradual accumulation of data from a variety of sources (Anastasi, 1982). There is construct validity in a critical thinking assessment tool if students who answer items correctly do so because of good critical thinking and not faulty reasoning, different levels of knowledge, or cultural differences (Arter, 1987; P. A. Facione, 1996). Research studies showing that critical thinking scores improve after training designed to improve critical thinking help determine its construct validity. There should also be a moderate correlation with intelligence and achievement tests (Arter, 1987).

### Critical Thinking Assessment Methods

Several methods may be used to evaluate and assess critical thinking. No one tool is adequate to measure the various dimensions of critical thinking (Maynard, 1996). Just as critical thinking is not developed by one method alone, it should not be evaluated by only one strategy (Oermann, 1997). Several authors call for multiple methods of assessment in a critical thinking assessment plan (N. C. Facione, 1995; Facione & Facione, 1994, 1996a; Videbeck, 1997b). Critical thinking assessment lends itself to a full array of methods including standardized commercial instruments and locally developed instruments with multiple choice questions, attitudinal inventories, essay tests, case study analysis, theoretical debates, role playing, and talk aloud exercises (N. C. Facione, 1995; Facione & Facione, 1994). There is an advantage to collect data from many areas to get a composite picture (P. A. Facione, 1996). Multiple measures also allow assessments in different contexts. The Commission on Institutions of Higher Education supports the use of a variety of approaches to document institutional

effectiveness such as senior seminars, senior recitals, capstone courses, surveys, activities with alumni, and locally developed exams (Payne, Vowell, & Black, 1991). External experts could be used to provide a non-biased assessment of programs, but they may not understand the uniqueness of the institutions as well as being expensive. Students may evaluate programs through surveys, inventories, and interviews but motivation to participate is low (Payne, Vowell, & Black, 1991). Student and graduate surveys are relatively inexpensive, easy to administer and can be related to local goals (Payne, Vowell, & Black, 1991).

Videbeck (1997a) proposes a model and process to give faculty direction in the creation of a plan to develop and assess critical thinking. This model addresses the dual purposes of evaluation of student achievement and program evaluation. In a descriptive study of current critical thinking assessment practices, Videbeck (1997b) found that 18 schools of nursing used only one measure of assessment and 37 used more than one. Twenty-nine programs used 44 standardized tests, 12 on entry and exit, 26 at the exit, and 2 after graduation. Twenty-two programs used locally developed assessment measures including graduate and employer surveys, a faculty developed clinical judgment tool, and case studies. Of the standardized tests, 9 schools used the Watson Glaser Critical Thinking Appraisal (WGCTA), 6 the NLN Diagnostic Readiness Test, 6 the NLN Comprehensive Test, 4 the California Critical Thinking Skills Test (CCTST), 3 the Mosby Assess Test, 1 the NLN Achievement Tests, 1 the California Critical Thinking Dispositions Inventory (CCTDI), 1 the ACT COMP, 1 the CAAP, and 1 the Cornell Critical Thinking Test (CCTT).

In a survey of BSN and MSN schools by O'Sullivan, Blevins-Stephens, Smith, and Vaughan-Wrobel (1997), 28% used standardized measures with the WGCTA and the CCTST being used most often, 2% used locally developed tests, 27% used individual assessments such as portfolios, written assignments, and clinical evaluations, 4% used surveys and 36% did not list any measures. Term papers and case studies were listed as ways to teach critical thinking as well as outcome measures making it difficult for students and faculty to differentiate the two processes.

Gendrop and Eisenhauer (1996) created an extensive list of suggestions for evaluating critical thinking in nursing. Included are diaries, innovative audits, summative and formative evaluations, portfolios, cognitive style appraisals (e.g. Kolb), essays, the CCTST, the CCTDI, the National Council Licensure Examination (NCLEX), and conceptual diagrams.

Alverno College, cited by many as exemplary in critical thinking outcomes assessment, defines assessment as a “multidimensional attempt to observe and, on the basis of criteria, to judge the individual learner in action” (Cromwell, 1992, p. 42). The college-wide plan includes a range of behaviors in multiple modes and contexts, incorporates open-ended methods, self-assessment, and structured feedback, and is cumulative and based on learning outcomes. The Alverno faculty believe that standardized objective tests cannot adequately measure a concept as complex as critical thinking (Cromwell, 1992; Halonen, 1995).

A major dichotomy in assessment methods is the use of standardized versus locally developed tools (Moss & Koziol, 1991). If the user wants useful information for monitoring achievement, then development, administration and scoring must be

standardized (Snyder, 1993). Commercial standardized tests use strict psychometric standards, provide consistent reliability, and can cover the critical thinking domain broadly. They are readily available without the delays, expenses and faculty time needed for locally developed tools, are generally easy to give and often machine scored, and can be compared to regional or national populations. They vary in the amount of feedback provided, may or may not be able to isolate areas of deficiency in the program, and are often voluntary, non-graded, and at student expense. They may also contain cultural, gender, racial, or geographic bias (Payne, Vowell, & Black, 1991). Locally developed measures, may not have the stringency of commercial tools, but can be more flexible and tailored to meet local needs (Ory, 1991). Although the value of standardized tools is debated, their use continues because of their availability and ease of use (Leppa, 1997).

The majority of the commercial tests are composed of multiple-choice questions. Advantages include ease of administration, speed of getting results, and provision of reasonably consistent results. Multiple-choice tests are useful for testing simple forms of critical thinking. The best method is the use of best answer with all choices correct and one slightly better than others (Haladyna, 1997). They are not adequate as the only measure of critical thinking. Critical thinking testing that requires one correct best answer is problematic because it reduces the validity of the tests as issues requiring critical thinking do not have a single correct answer (Norris, 1988).

Another area of concern is that multiple-choice questions focus on product rather than process. Since critical thinking is difficult to measure directly, tools often make inferences from the product of student's thinking to the construct (Norris & Ennis, 1989). Multiple-choice questions do not give any indication of the thinking processes that led to



the answer choice. This method is not suited for complex problems where students need to raise questions or use several abilities together (Moss & Koziol, 1991). Neither is it a useful measure for assessing dispositions. Multiple-choice tests provide weak evidence of thinking processes but there is better support if verbal reports or written justification of thinking choices and rationale are used (Norris, 1990). If written justification is included, answers that differ from the key could be given credit for valid reasoning. This adjustment allows for differences in student backgrounds (Ennis, 1993). Using verbal reports also helps to increase the validity of the tool by showing that correct answers have been arrived at via good critical thinking. Methods for getting verbal reports include think aloud, immediate recall, and criteria probe. A study evaluating these methods as well as a control group of no elicitation found that these methods did not alter critical thinking performance but there were some interviewer effects (Norris, 1990). There are four criteria for writing critical thinking multiple choice questions which include having written rationale, being written at the application level or higher, requiring multilogical thinking, and requiring a high level of discrimination (Morrison, Smith, & Britt, 1996). Multilogical thinking requires considering and reasoning within multiple points of view (Morrison, Smith, & Britt, 1996).

As an alternative to multiple-choice tests, constructed response questions allow for more than one legitimate answer to a problem. Students have an opportunity to demonstrate they can use several critical thinking abilities on a complex problem. Critical thinking dispositions can also be measured. They offer more clearly valid information on critical thinking abilities than multiple-choice tests. Disadvantages include difficult and time consuming scoring (Norris & Ennis, 1989). They tend to have

trouble with lower inter-rater reliability and difficulty covering a wide range of critical thinking abilities and dispositions in a reasonable time (Halpern, 1993; Norris, 1988). Until there is a greater number of open-ended critical thinking tests, faculty may want to make their own open-ended tools for a comprehensive assessment (Ennis, 1993).

There are several alternatives to commercial standardized instruments. Moss and Koziol (1991) expressed a concern that traditional standardized assessment tools have lowered academic standards to the level that can be easily and reliably evaluated. A 1990 survey showed use of local instruments increased from 55% in 1989 to 66% in 1990 (Ory, 1991). Several experts suggest that locally developed measures may conform to a program's goals more than commercially developed tests (Rane-Szostak & Robertson, 1996; Videbeck, 1997a). Developing local instruments may be expensive. There is a trend to open-ended assessment such as portfolios, interview and focus groups (Halpern, 1993). Oermann (1997) includes methods for evaluating critical thinking in the clinical setting such as observing with questions about rationale, Socratic questioning, critical incident conferences, and written assignments.

Portfolios allow documentation of growth in critical thinking over time and are not very expensive in terms of time and cost (Rane-Szostak & Robertson, 1996). The primary concern about these and other related techniques is interrater reliability (Halpern, 1993). An evaluation tool such as the Holistic Critical Thinking Scoring Rubric will help improve interrater reliability (Facione & Facione, 1996b; Rane-Szostak & Robertson, 1996).

Martin, Kinick, Hummel, Clukey, and Baird (1997) describe a portfolio process piloted in a nursing program. It proved to be an excellent measure for evaluating

individual student performance, but was not helpful in providing aggregate data for evaluation of the program. Instead, the faculty developed an Outcome Evaluation Tool and an Outcome Essay with rating criteria that were used at the end of the program to produce aggregate means for each criterion that faculty can use to evaluate the program.

Performance assessment involves observation of clinical effectiveness. This method is expensive as it requires considerable time (Ennis, 1993). Whitney (1993) reports that performance assessment needs more work on validity, reliability, fairness, practicality and efficiency before it can be used a part of a national assessment policy. Technology such as use of videos or computer simulations could be used to improve its efficiency. Results should be used for program or group evaluation, and not for making decisions about individual students (Whitney, 1993). Performance assessment used to be the method of choice in critical thinking assessment before the switch to essays and then the current reliance on multiple-choice formats. Whitney (1993) predicts a reversal back to increased use of performance assessment in the future. Regents College in New York is an example of the use of performance assessment in nursing (Rane-Szostak & Robertson, 1996). It is based on assessment of knowledge and competence and not instruction (Lenburg & Mitchell, 1991).

In summary, the literature documents advantages and disadvantages of each of the different methods of assessing critical thinking. A critical thinking assessment plan should utilize several different methods, including multiple choice and open-ended formats with standardized and locally developed instruments.

## Strategies for Improving Critical Thinking in Nursing

Because the ultimate goal of outcomes assessment is the improvement of programs, there is an emphasis in the literature on ways to enhance critical thinking in nursing. Alfaro-LeFevre (1995) presents a long list of general strategies to improve critical thinking, including anticipating questions from others, asking why, what else, and what if questions, paraphrasing, comparing and contrasting, organizing and reorganizing information, looking for flaws in thinking, and turning errors into learning opportunities. Additional questions to ask are what do we know, how do we know, what do we accept, and what is the evidence for (Schank, 1990). Carlson-Catalano (1992) discusses empowering strategies of analytic nursing, change activities, collegiality and sponsorship as ways to improve critical thinking. Videbeck (1997a) stresses increased use of varied classroom strategies and increased use of collaborative teaching techniques. Dobrzykowski (1994) used Benner's novice to expert research to address strategies to enhance critical thinking in the practicing nurse including mentoring activities such as thinking aloud and debriefing, chart review, critical incident discussions, mindmapping, computer simulations and videotapes. Similarly, Kennison and Brace (1997) suggest listening to experienced nurses problem solve as a strategy, and Paul (1992) encourages faculty to think aloud in front of students. Wink (1993) discusses the use of convergent and divergent questioning, and Paul (1992) suggests questioning students Socratically. Based on John Dewey's philosophy of inquiry, the Personally Perceived Problem Technique fosters critical and creative thinking by exploring the situation, idea generation, solution validation and evaluation (Russaw, 1997). Baker includes identifying assumptions, fallacies in reasoning, rival causes, omitted information, value conflicts,

reasons, and strength of evidence (Beeken, Dale, Enos, & Yarbrough, 1997). Although not reported as a research study, the author concludes that there is evidence that critical thinking development may be improved by well-designed course content on critical thinking since the mean scores on the California Critical Thinking Skills Test were 16.72 and 17.71 compared to a standardized norm of 15.89. It is not reported whether the results are significant or not, but the author suggests that a different tool, the California Critical Thinking Dispositions Inventory may be more appropriate because the real value of the course is enhancement of critical thinking dispositions (Beeken, Dale, Enos, & Yarbrough, 1997).

More specifically, Durgahee (1996) endorses the use of a reflective diary to enhance reflective reasoning. The author used survey and semi-structured interviews in a qualitative study to show that use of reflective diaries led to increased intellectual focusing and questioning of practices and resulted in critical thinking. Baker (1996), Hahnemann (1986), Farrell (1996), Dobrzykowski, (1994), and Brown and Sorrell (1993) also discuss use of structured reflective clinical journal writing to increase critical thinking skills. Participation learning activities such as using costumes to portray nursing diagnoses is suggested as another way to help students develop reflective practice (Sedlak & Ludwick, 1996).

The use of case studies to identify and challenge assumptions, and explore alternative ways of thinking is included in almost every list of critical thinking strategies (Calderone, 1997, Chubinski, 1996; Farrell, 1996; Morrison, Smith, & Britt, 1996; Neil, Lachat, & Taylor-Panek, 1997; Schumacher & Severson, 1996; Vanetzian & Corrigan, 1996). Schumacher and Severson (1996) shared a collaborative education practice project

called “Critical Thinking Rounds” which uses a case study approach to enhance critical thinking development. Neil, Lachat, and Taylor-Panek (1997) note that the impact of the case study approach is difficult to measure. Their program uses course evaluations to evaluate this strategy but the tool has not been tested for reliability and validity. White, Beardslee, Peters, & Supples (1990) add argumentation, debate, analyzing and interpreting ethnography, and ethical decision making. Creative teaching strategies such as an intraoperative awareness exercise, and ER/ICU pharmaceutical decision making are suggested as ways to improve critical thinking in pharmacology (Wissmann, 1996). The AIDS Care Dilemma Exercise, a written paper and class discussion, is a strategy to develop critical thinking by challenging the students to consider alternative perspectives (Lewis & Eakes, 1992). A community health project helps students explore values and assumptions, recognize the importance of context, and explore alternate solutions (Kuennen & Moss, 1995). Another specific strategy includes the use of jigsaw, a versatile cooperative learning strategy that calls for use of analysis, reconstruction, reflection and synthesis (Ulrich & Glendon, 1995). Kurfiss adds using controversy, student to student dialogue, simulations, and critiquing research (cited in Schank, 1990). Another list includes summarizing, games, role-playing, flow diagrams, cognitive maps, and information organizers (Calderone, 1997). An interdisciplinary health issues course focused on providing contextual variation and writing-to-learn strategies to improve cognitive skill development in pre-nursing students (Bowers & McCarthy, 1993). Elliott (1996) notes that most critical thinking enhancing strategies, such as case studies, are better suited for small groups and instead lists strategies for large classes such as making learning meaningful, creating disequilibrium, thinking pair-share, round robin, creating a

jigsaw, and allowing note monitoring. Thompson and Rebesch (1998) add concept mapping, examining research findings and identifying improvements for nursing practice, identifying similarities and differences between approaches, questioning the credibility of others, and using various theoretical frameworks to explain and predict client behavior.

Gendrop and Einsenhauer (1996) have created an extensive list of suggestions for teaching-learning strategies for critical thinking in nursing. Examples include active involvement strategies such as role-playing and debate, reflection strategies such as diaries and ethnography, self-disclosure activities such as critical thinking dispositions and skill appraisals, collaborative reflective strategies such as patient-family-nurse dialogue and multicultural dialogue, cognitive strategies such as brainstorming, contextual strategies such as case studies, and noncontextual strategies such as imagery.

Several non-nursing authors suggest strategies for improving critical thinking. Angelo (1995) inferred from the research of McKeachie, Pintrich, Lin, and Smith that student discussion, explicit emphasis on problem solving and verbalization of metacognitive strategies enhance critical thinking. Cross notes that classroom assessment techniques tied to instruction are a useful tool (Angelo, 1995). Lantz (1986) and Lashley and Wittstadt (1993) see writing activities as a method for enhancing critical thinking across the curriculum. These methods include journal writing, personification, portfolios, and other assignments that build on one another, reflect increasing complexity, and encourage application of knowledge and concepts in different contexts. Kramer (1993) and Kemp (1985) suggest use of concept clarification as a critical thinking enhancing strategy. Pascarella and Terenzini performed a study that showed that gains in critical thinking were related to three teaching behaviors, faculty encouragement and use of

student ideas, level of student participation in class, and peer-to-peer interaction (cited in Cross & Steadman, 1996).

Gokhale (1995) used an instructor-developed test to measure critical thinking in an industrial technology course in a study to evaluate the effectiveness of individual learning vs. collaborative learning in enhancing critical thinking skills. No significant difference was found between groups on lower order thinking skills, but the collaborative learning group scored significantly higher on the critical thinking items (Gokhale, 1995). P. A. Facione (1991) and Nummedal (1991) have called for criteria for assessment of critical thinking in a group context since everyday problem solving often occurs in interaction with others.

Halpern and Nummedal (1995) list general strategies for improving critical thinking such as testing hypotheses, making decisions, supporting conclusions with evidence, maintaining an open mind and working collaboratively. Heyman and Daly (1992) describe the use of giving partial credit for reasoning used in problem solving in occupational programs. Paul (1985) supports giving students practice in reasoning dialogically and dialectically so they get skilled in weighing, reconciling, and assessing contradictory points of view. Kumar and Fritzer (1994) advocate an interdisciplinary approach to educating for critical thinking which helps students better develop problem solving and decision making skills by asking questions that deal with more than one subject.

Most of the strategies suggested to improve critical thinking are presented on faith that they do indeed improve critical thinking. None of the specific nursing strategies had their effectiveness supported by research. The only evaluation method mentioned in the



nursing literature was end of course evaluations. Research is needed to determine the effectiveness of each of the strategies listed.

## Standardized Critical Thinking Assessment Tools and Associated Research Studies

### The Watson Glaser Critical Thinking Appraisal

The Watson Glaser Critical Thinking Appraisal (WGCTA), first developed in the 1940s and revised in 1980, has been around the longest and is the benchmark against which other tools have been measured (Facione & Facione, 1994). It has been one of the tools utilized most often for research and evaluation (Miller, 1992). Written for 9<sup>th</sup> grade through adult, it is a general skills test using everyday work problems that takes 40 minutes. It consists of 80 general items on 2 forms divided into five subtests: inference, recognition of assumptions, deduction, interpretation, and argument evaluation (Carpenter & Doig, 1998). The two forms, A and B, are balanced for difficulty, content, correlation with total score and are considered equal and alternate forms (Adams, Whitlow, Stover, & Johnson, 1996). This tool was originally developed as a measure of the dependent variable in an experiment designed to assess the effects of teaching high school students critical thinking (Jacobs, 1995). It was based on the five abilities of defining a problem, selecting significant information for solution of problems, recognition of assumptions, formulation and selection of hypotheses, validation of conclusions and judgment of inferences (Adams, Whitlow, Stover, & Johnson, 1996). Although it does not test credibility of sources and observations or critical thinking dispositions, it is considered balanced and comprehensive (Norris & Ennis, 1989). Split half coefficients range from .69 to .85 (Adams, Whitlow, Stover, & Johnson, 1996; Rane-

Szostak & Robertson, 1996). Test-retest reliability is reported as .73 and alternative form reliability is .75 (Adams, Whitlow, Stover, & Johnson, 1996). Helmstadter (1985) notes that reliability is not as high as is preferred, but this may be because four of the five subtests are composed of items with only two alternatives, or because of judgmental components in the inferential subtest. Reliability coefficients for the five subtests are .41 to .74. Because of the low reliabilities for subtests, Watson and Glaser recommend the subtests be used to analyze critical thinking ability of groups rather than individuals (Miller, 1992). Validity is supported by studies that showed increases in scores following critical thinking instruction and correlation with measures of general intelligence, aptitude and achievement (Norris & Ennis, 1989). A panel of twelve psychologists published in critical thinking evaluated the WGCTA a superior tool to the Cornell Critical Thinking Test which is discussed later in this chapter (Jacobs, 1995). This evaluation was based on the degree to which it met the 1974 Standards for Educational and Psychological Tests. These standards include ten essential validity standards, and five essential reliability and measurement error standards. The WGCTA met the criteria for 9 standards. It had an unfavorable evaluation for possible test bias and lack of cross validation efforts and was found to need revision to enhance reliability and validity (Modjeski & Michael, 1983). Several reviewers expressed caution about the use of scores from the WGCTA, but still recommend it due to the availability of equivalent forms, extensive normative data, and other technical data (Jacobs, 1995). Correlations between subtests show substantial relationships in the .56 to .79 level. The median score for college students is in the low 52 to 60 range (Norris, 1985). A. Berger (1985) labeled the WGCTA as a well-constructed test. Ennis (1993) listed its weaknesses as its narrow

scope and high dependence on critical thinking definition. Limitations of the WGCTA have been noted to be in construct validity, normalization data, assessment of inferences as true/false rather than valid/invalid, and appraisal through reading (Pless & Clayton, 1993). Davis (1991) summarized the characteristics of the WGCTA as having good norms, fair to good reliability, and extensive validity.

There is also a WGCTA Form S, which is a shorter, updated version of Form A with 40 items to be given in 30 minutes. The same subtests are used but the scores should not be used for individual use. A concern was raised by a reviewer that 95% of the norm group is spaced over only 15 points, causing a large amount of uncertainty in scores. A single question in some norm groups could make a major change in percentile ranking (Geisinger, 1998).

The majority of the studies related to critical thinking in the nursing literature have used the WGCTA (Adams, Whitlow, Stover, & Johnson, 1996). Table 1 summarizes results of nursing studies using the WGCTA. No studies using the short form were found. Longitudinal and cross-sectional studies looking at the effect of nursing education on critical thinking abilities using the WGCTA show inconsistent results (Kintgen-Andrews, 1991). Research results reported in the education literature are no more conclusive than nursing (Kintgen-Andrews, 1991). Longitudinal studies expect improvement from entry to exit because of the emphasis on the nursing process and the utilization of the concept of critical thinking as a problem solving process (Gross, Takazawa, & Rose, 1987). For BSN students, the pre-test is usually given at the beginning of the junior year or at the end of the sophomore year, and the exit test is given

Table 1 – Summary of Nursing Studies on Critical Thinking Using the Watson Glaser Critical Thinking Appraisal

Researcher	Type of Study	Sample	Major Findings about Critical Thinking	Variable showing relationship to critical thinking	Variables showing no relationship to critical thinking
Frederickson, 1979	Longitudinal	14 BSN	Significant increase from entry to exit	GPA	
Berger, 1984	Longitudinal	137 BSN	Significant increase from entry to exit		Nursing & Science GPA
Bauwens & Gerhard, 1987	Longitudinal	159 BSN	No significant increase from entry to exit	Graduate GPA, Nursing GPA, Entry GPA, NCLEX-RN passage	
Gross, Takazawa & Rose, 1987	Longitudinal	37 ADN 45 BSN	Significant increase from entry to exit	(-) GPA for ADN & (+)GPA for BSN on exit NCLEX-RN for BSN	
Sullivan, 1987	Longitudinal	46 RN-BSN	No significant differences from entry to exit		
Kintgen-Andrews, 1988	Longitudinal	55 VN 55 ADN 38 Pre-health science freshmen 29 sophomore BSN	No groups made significant gain over one year		
Miller, 1992	Longitudinal	137 RN-BSN	Significant increase from entry to exit for diploma, not ADN	Nursing GPA	

Researcher	Type of Study	Sample	Major Findings about Critical Thinking	Variable showing relationship to critical thinking	Variables showing no relationship to critical thinking
Saucier, 1995	Longitudinal	97 BSN 31 RN-BSN	No significant increase for each group over 3 years except 1 BSN group		
Behrens, 1996	Longitudinal	41 Diploma	No significant increase from entry to exit	GPA, Age	
Maynard, 1996	Longitudinal	24 RNs (BSN prepared)	No significance increase from sophomore to senior but significant increase from senior to practice		
Nathan, 1997	Longitudinal	2 classes	No significant increase over 1 year		GPA, learning style
Vaughan-Wrobel, O'Sullivan & Smith, 1997	Longitudinal	391 BSN	No significant increase for 3 classes from entry to exit	Age, Completion of another degree	Previous nursing experience
Whiteside, 1997	Longitudinal	27 RNs	Significant increase after 10 week critical care course using model for critical thinking		
Frye, Alfred, & Campbell, 1999	Longitudinal	27 BSN	No significant difference from freshmen to senior level		

Researcher	Type of Study	Sample	Major Findings about Critical Thinking	Variable showing relationship to critical thinking	Variables showing no relationship to critical thinking
Richards, 1977	Posttest only	217 BSN (one group - block; two groups-integrated)	Students in new integrated curriculum had significantly lower scores than those in old block curriculum		
Hartley & Aukamp, 1994	Posttest only	50 nursing faculty	Faculty scored significantly higher compared to nursing student norms		
Perciful & Nester, 1996	Posttest only quasiexperimental	44 BSN – control group 39 BSN exp. group	Significant higher scores for experimental group		
Frederickson & Mayer, 1977	Cross-sectional	28 BSN 27 ADN	BSN scored significantly higher than ADN		
Polifroni, 1981	Cross-sectional	90 pre-BSN 85 BSN 83 RN-BSN	Pre-nursing and BSN scored significantly higher than RN-BSN		
Scoloveno, 1981	Cross-sectional	90 BSN 93 ADN 97 Diploma	BSN students scored significantly higher than ADN and Diploma		
Pardue, 1987	Cross-sectional	121 RNs (24 Diploma, 27 ADN, 33 BSN, 37 MSN)	Significant difference across levels with BSN and MSN highest		

Researcher	Type of Study	Sample	Major Findings about Critical Thinking	Variable showing relationship to critical thinking	Variables showing no relationship to critical thinking
Brigham, 1989	Cross-sectional	114 BSN (4 levels)	No significant differences in 4 levels	SAT verbal and quantitative, GPA, Age, Total # of credit hours	
Lynch, 1989	Cross-sectional	74 BSN 87 ADN	BSN scored significantly higher than ADN	SAT scores, educational level	Age
Kokinda, 1990	Cross-sectional	49 BSN (4 levels)	Significant differences over 4 levels in inference, deduction, and evaluation of arguments	Academic achievement	
Brooks & Shepherd, 1990, 1992	Cross-sectional	50 ADN 50 Diploma 50 BSN 50 RN-BSN	BSN & RN-BSN scored significantly higher than ADN & Diploma	Age for RN-BSN	
Saarman, Freitas, Rapps, & Riegel, 1992	Cross-sectional	32 sophomores 32 ADN 32 BSN 32 Faculty	No significant difference between groups when adjusted for influence of age	Age	
Bingaman, 1993	Cross-sectional	89 ADN 28 RNs	Significant difference between RNs and all student groups		
Sietsema, 1993	Cross-sectional	19 BSN 15 ADN	No differences between groups	GPA	

Researcher	Type of Study	Sample	Major Findings about Critical Thinking	Variable showing relationship to critical thinking	Variables showing no relationship to critical thinking
Smith, 1995	Cross-sectional	239 ADN	1 <sup>st</sup> year & low academic achievers proficient in lower cognitive skills; 2 <sup>nd</sup> year & high academic achievers higher in high cognitive skills	GPA; reading and math abilities; learning and study strategies	
Stadt, 1995	Cross-sectional	357 Certificate, AD, and BS radiology students	BS significantly higher than AD & Certificate; Certificate significantly higher than AD		Age, gender, education completed, work experience
Clocklin, 1996	Cross-sectional	197 1 <sup>st</sup> year nursing students from LPN, ADN and BSN	Significant relationship between critical thinking skills and preferred learning style	Age	
Howenstein, Bilodeau, Brogna, & Good, 1996	Cross-sectional	152 RNs (18 diploma, 5 ADN, 70 BSN, 58 MSN, 1 PhD)	BSN & MSN scored significantly higher than ADN & Diploma	Age (-), Years of Experience (-), Level of education (+)	Practice area
Kelly, 1996	Cross-sectional	RNs	Baccalaureate education had significant effect on critical thinking ability	Clinical experience	
Frye, Alfred, & Campbell, 1999	Cross-sectional	132 Freshmen & 77 Seniors - BSN	Significant difference between 2 levels		



Researcher	Type of Study	Sample	Major Findings about Critical Thinking	Variable showing relationship to critical thinking	Variables showing no relationship to critical thinking
Tanner, 1977	Correlational	54 BSN	No relationship between critical thinking and 7 measures of clinical judgment		
Ketefian, 1981	Correlational	79 RNs	The higher the critical thinking scores, the higher the moral reasoning		
Gunning, 1982	Correlational	66 BSN	Significant relationship with field independence; no significant relationship with clinical problem solving		
Holzemer & McLaughlin, 1988	Correlational	34 Nurse Practitioners	No significant correlation with simulation tests		
Ircink Waite, 1989	Correlational	299 Senior BSN	No significant relationship between critical thinking and 11 different curricula models	GPA	Age, sex, years of work experience, education

at graduation. Students have completed two years of general education courses before the pre-test.

The great majority of these studies used BSN students. In the earliest longitudinal study using the WGCTA, Frederickson studied a small sample of 14 BSN students in a pilot study and reported that critical thinking abilities improved from entry to exit (cited in Adams, Whitlow, Stover, & Johnson, 1996). Frederickson also found a relationship between posttest critical thinking scores and grade point average (GPA) (Miller & Malcolm, 1990).

M. C. Berger (1984) also cited statistically significant improvement in critical thinking scores on the WGCTA for BSN students from entry to exit, providing support for the belief that creative and critical thinking can be improved by the student faculty relationship and the learning environment. No significant relationship was found between critical thinking scores and nursing or science GPA.

Bauwens and Gerhard (1987) found no significant difference in critical thinking scores of BSN students from entry to exit. The authors suggested that the WGCTA may be more useful as an admission screening tool for applicants than as a predictor of nursing success.

In one of the first studies that included ADN students, Gross, Takazawa and Rose (1987) found highly significant improvement in mean critical thinking scores on the WGCTA for ADN and BSN groups from entry to exit. Both groups had similar scores on entry. For the BSN group, critical thinking was a predictor of performance on the NCLEX-RN. The licensing exam scores were not predictable by critical thinking scores at exit, suggesting that the exam tests critical thinking less than academic performance

(Gross, Takazawa, & Rose, 1987). The authors reported a significant positive relationship between GPA and critical thinking for BSN, but a negative relationship for ADN students (Vaughan-Wrobel, O'Sullivan, & Smith, 1997).

Sullivan (1987) found no relationship between clinical skills and WGCTA scores and no difference between entry and exit scores on the WGCTA for BSN completers. BSN completers are associate degree or diploma prepared registered nurses with varying experience returning to get their baccalaureate degree. Kintgen-Andrews (1988) found no significant gain in critical thinking scores over one year with LVN, university pre-health science freshmen, ADN and BSN sophomores.

Miller (1992) found a significant relationship between nursing GPA and exit WGCTA scores, and a significant difference between entry and exit WGCTA scores for BSN completers in a one group pre-test/post-test ex-post facto design. Both ADN and BSN graduates showed an increase in low cognitive domain thinking but no difference in the high cognitive domain. In this study, posttest means were at the 41.25 percentile rank compared to college senior women. Examination of the statistics showed that the only gains were found in diploma graduates and not in associate degree graduates. Most change occurred in the recognition of assumptions and deduction subtests. The author suggested that the ADN graduates did not show the same gains as the diploma graduates because they had already been exposed to two years of higher education (Miller, 1992). Saucier (1995) found no significant difference in critical thinking ability of three classes of RN to BSN students and two classes of generic BSN students as measured by the WGCTA, and a significant difference in only one class of generic BSN students from entry to exit.

Behrens (1996) studied nursing students in diploma programs (three-year hospital-based nursing programs) and found critical thinking scores on the WGCTA did not increase from entry to exit. However, WGCTA scores were found to be a predictor of nursing school outcome. A positive relationship was found between chronological age, critical thinking ability and GPA.

Maynard (1996) studied a randomly selected cross sectional sample of BSN graduates and found that there was no significant change in critical thinking scores from sophomore to senior level, but there was a significant change from senior to practicing nurse. Critical thinking abilities had no effect on measures of nursing competence, but there was a highly significant correlation between level of competence and years of practice. There was no significant relationship between their ability to derive nursing diagnoses and critical thinking ability on the WGCTA (Maynard, 1996).

Bechtel, Smith, Printz, and Gronseth found no support for the impact of nursing education on critical thinking (cited in Gendrop & Eisenhauer, 1996). Nathan (1997) found no significant difference between critical thinking ability of two BSN classes over one year. Vaughn-Wrobel, O'Sullivan and Smith (1997) used the WGCTA with BSN students on entry and exit, with an additional measurement at the end of the junior year, and found a positive correlation between higher critical thinking scores, age, and completion of a previous degree. After controlling for those variables, no significant differences were found from entry to exit.

Whiteside (1997) found significant improvement in critical thinking scores on the WGCTA of an experimental group of students when faculty used a model for teaching critical thinking in the clinical setting over 10 weeks. In the latest longitudinal study

found in the literature, Frye, Alfred, and Campbell (1999) found no significant differences in WGCTA scores in a group of 27 BSN students from their freshman year to senior year.

In one of the earliest nursing studies of the impact of nursing education on critical thinking, Richards (1977) used a post test only format to test the effect of a new integrated BSN curriculum vs. the previous block curriculum on critical thinking ability as measured by the WGCTA. They found that the integrated curriculum students had lower critical thinking scores at exit. This was attributed to the focus on one way to solve problems in the integrated curriculum.

Another post-test only study investigated the critical thinking abilities of nurse educators. Hartley and Aukamp (1994) compared critical thinking scores of a small sample of educators to norms of BSN students in the south, west and midwest. They found that faculty had significantly higher means of critical thinking ability.

Percival and Nester (1996) conducted a quasi-experimental study to determine the effect of an innovative clinical teaching method with CAI and a collaborative student-faculty-service model on knowledge and critical thinking skills in BSN students in a psychiatric clinical. The scores on assessment, analysis, and evaluation on a NLN psychiatric exam and another score for planning and implementing on the same exam were used as measures of critical thinking. Students in the experimental group scored higher on both critical thinking measures but there was no difference in knowledge level. The assessment, analysis, and evaluation scores had a stronger correlation with the WGCTA scores. Parts of the nursing process are not appropriate indirect measures of

critical thinking. This study was limited by sample size and setting (Perciful & Nester, 1996).

With cross-sectional research, the abilities of students prior to entering the program are in question. In addition, admission criteria must be considered since the progressive selectivity of students in BSN programs may affect the results (Hickman, 1993, Kintgen-Andrews, 1991). Most of the cross-sectional studies tested the differences found between ADN and BSN students or between different levels of BSN students. The results were also mixed. Several showed no significant differences among groups but most found that higher scores correlated with higher levels of education.

Frederickson and Mayer (1977) studied critical thinking ability using the WGCTA. They found that BSN students performed significantly higher than ADN students on the critical thinking tool, but not on a problem-solving tool. Matthews and Gaul (1979) found no significant difference between graduate students and undergraduate students on the WGCTA but graduate students performed better on clinical judgment.

Polifroni (1981) noted that pre-BSN and BSN students scored higher in critical thinking than RN-BSN completers. Scoloveno (1981) found BSN graduates had significantly higher critical thinking scores on the WGCTA and the Nursing Process Utilization Inventory than ADN graduates.

Pardue (1987) found a significant difference in critical thinking abilities between all four levels of preparation (ADN, Diploma, BSN, and MSN) in a sample of experienced practicing nurses. The BSN and MSN graduates had the highest mean. There was also no effect of education or performance on clinical judgment (Hickman, 1993; Miller & Malcolm, 1990). Brigham (1989) showed no significant difference between

BSN students in a stratified random sample from each class from freshman through senior levels. Lynch (1989) found BSN graduates' scores significantly higher than ADN graduates on the WGCTA.

Kokinda (1990) found significant differences in subtest scores of a stratified random sample of four levels of generic BSN students for inference, deduction, and evaluation of arguments, but not for recognition of assumptions and interpretation.

Brooks and Shepherd (1990, 1992) studied critical thinking, clinical decision making, and professionalism across four types of nursing programs with a convenience sample of ADN, BSN, Diploma, and RN completer volunteers. Results showed that BSN students had higher critical thinking ability than ADN or diploma students did. The higher scores of RN-BSN graduates suggest experience is a variable affecting critical thinking scores (Hickman, 1993). The authors also found a weak but significant positive relationship between critical thinking and clinical decision-making, as measured by the Nursing Performance Simulation Instrument and the Clinical Decision Making in Nursing Instrument (Brooks & Shepherd, 1990). Clinical decision making was highest for RN completers (Brooks & Shepherd, 1992). There was no significant difference between levels of professionalism of BSN and ADN students (Brooks & Shepherd, 1992).

A cross-sectional correlation survey by Saarman, Freitas, Rapps, and Riegel (1992) found no significant difference between critical thinking ability of a convenience sample of faculty, BSN sophomore and senior students, and ADN senior students when age was controlled statistically. Based on social learning theory it was thought that students exposed to faculty in longer curricula would have higher critical thinking levels.

The assumption is that faculty have higher critical thinking skills. The wide range of ages and years of nursing experience were cited as limitations.

Bingaman (1993) found statistically significant differences in mean critical thinking scores on the WGCTA between RNs and all ADN student groups. Sietsema (1993) found no significant differences on the WGCTA for ADN and BSN seniors.

M. L. Smith (1995) studied 239 ADN students and found a Pearson correlation coefficient of .4138 between GPA and Watson Glaser Critical Thinking Appraisal scores. First year students and low academic proficiency students were proficient in the low cognitive critical thinking abilities such as deduction, and second year students and high academic proficiency students had higher scores on interpretation and evaluation of arguments, the higher critical thinking abilities.

In a study using the WGCTA and students in baccalaureate, associate degree, and certificate radiology programs, baccalaureate students scored significantly higher than associate degree and certificate students and certificate students scored significantly higher than associate degree students. Age, gender, education, and work experience were not seen as having significant effects on the critical thinking scores (Stadt, 1995).

Clocklin (1996) found a significant relationship between critical thinking skills and preferred learning styles of a group of LPN, ADN, and BSN students. Convergers had the highest scores. Age was related to critical thinking scores.

In an exploratory, cross-sectional survey of a convenience sample of practicing nurses, Howenstein, Bilodeau, Brogna, and Good (1996) found that age and years of experience negatively correlated with critical thinking. The level of education made a



significant difference in critical thinking scores with BSN and MSN graduates scoring higher, but only five nurses in the study were ADN prepared.

Kelly (1996) found clinical experience and baccalaureate education had a positive effect on critical thinking abilities of registered nurses on the WGCTA with clinical experience having the greatest impact. Bell found no significant difference between critical thinking abilities of ADN and BSN students using the WGCTA (cited in Colucciello, 1997). Frye, Alfred, and Campbell (1999) found a significant difference between WGCTA scores of a cross-sectional sample of freshmen and senior BSN students even though they did not find a significant difference in their longitudinal sample.

Several correlational studies were reported in the literature. Studies focusing on correlation with clinical judgment were contradictory, but overall showed positive effects of nursing education on clinical judgment but no correlation between clinical judgment and critical thinking (Hickman, 1993; Kintgen-Andrews, 1991). In 1977, Tanner found no significant relationship between critical thinking and seven dependent measures of clinical judgment (cited in Miller & Malcolm, 1990; Pless & Clayton, 1993). Ketefian (1981) found the higher the critical thinking abilities of practicing nurses, the higher their moral judgment and reasoning scores. The BSN prepared nurses had higher levels of moral reasoning than ADN or diploma-prepared nurses. Adams, Whitlow, Stover, and Johnson (1996) noted that no evidence of congruence between critical thinking and clinical judgment was found.

Several other variables have been correlated to critical thinking scores on the WGCTA. Gunning (1982) found no relationship between critical thinking ability and

clinical problem solving on the Nursing Performance Simulation among BSN seniors. Holzemer and McLaughlin (1988) found no correlation between clinical simulation scores and critical thinking for nurse practitioners. Bauwens and Gerhardt (1987) found a .31 correlation between WGCTA critical thinking scores and NCLEX-RN scores. Kintgen-Andrews (1988) found a correlation of .24 to .65 between WGCTA and GPA, and a .34 to .56 correlation between WGCTA and the NCLEX-RN.

Ircink Waite (1989) found no significant relationship between three curriculum models of BSN programs and critical thinking scores on the WGCTA and no significant relationship among age, sex, years of work experience or education and critical thinking scores. There was a significant positive relationship between critical thinking scores and GPA.

Studies using the WGCTA in a California community college campus found 61% below the 50<sup>th</sup> percentile with significant positive correlation between critical thinking skills, age, and completion of college hours (Pearson, 1991). A significant relationship was found between critical thinking and ethnicity with a greater percentage of minorities scoring below the 50<sup>th</sup> percentile.

The broad nature of WGCTA may minimize the ability of students' critical thinking scores to be affected by a single course or teaching strategies. It is difficult to measure changes in a broad concept that is influenced by many factors over a long period of time, by changing only one factor (Pless & Clayton, 1993). It is also difficult to measure a complex construct using a broad objective test (Pless & Clayton, 1993; Vaughan-Wrobel, O'Sullivan, & Smith, 1997). Two reviews on critical thinking studies in nursing found the majority use the WGCTA with 16 of 27 studies in one review in

1987 (Pless & Clayton, 1993) and 21 out of 25 studies in the other review in 1990 (Vaughan-Wrobel, O'Sullivan, & Smith, 1997). Nursing tended to use this tool more often because it was most closely aligned with the nursing process. Miller and Malcolm (1990) suggest that the WGCTA is the most useful tool for nursing because of the practice focus of nursing education. Several researchers suggest that the WGCTA is not appropriate for measuring thinking skills developed through the nursing process (Maynard, 1996). It has been suggested that many studies show no change because the WGCTA puts an emphasis on logic and not process. The expectation of gain in critical thinking scores in upper division students may be flawed as most significant gains have been reported in the freshman year (Vaughan-Wrobel, O'Sullivan, & Smith, 1997). Another hypothesis is that nursing curricula may not be designed to enhance critical thinking (Mackie & Graham, 1996; Vaughan-Wrobel, O'Sullivan, & Smith, 1997).

Several reviews rated the WGCTA as a better tool than others available at the time it was revised in 1980, even though its reliability was not as high as preferred, especially with the subscales. The WGCTA has been used most often in studies of critical thinking in nursing. These studies tend to be divided into two major groups, longitudinal, measuring critical thinking from entry to exit with mostly BSN students, and cross-sectional comparing different levels of nursing students. None of the longitudinal studies used multiple comparison groups including those not in college to separate out maturational gains or gains from the college experience itself. The majority of the research conducted using the WGCTA has been with one school and convenience samples. The studies by Brigham and Maynard used randomly selected samples, and neither of these found significant differences. Concerns about the inconsistent results in

the field of nursing have led researchers to investigate other more recently published tools. In addition, the possible mismatch of the definition of critical thinking on which the tool is based and current thinking raises questions about the usefulness of the tool.

### The Cornell Critical Thinking Test

Ennis, Millman and Tomko created the Cornell Critical Thinking Test (CCTT) (Level Z) for gifted high school and college students and adults in 1985 (Jacobs, 1994; Carpenter & Doig, 1998). It was originally designed to evaluate the effectiveness of critical thinking instruction and was the result of long term research (Jacobs, 1995). Fifty-two items measuring induction, deduction, observation, credibility, definition, predictive and experimental planning, fallacies, and assumption identification take 50 minutes to administer (Adams, Whitlow, Stover, & Johnson, 1996; Carpenter & Doig, 1998; Ennis, 1993). This philosophically derived tool was based on Ennis' definition of critical thinking. The median scores for undergraduate university students is 30 out of 52, suggesting that critical thinking scores are not high at any level (Norris, 1985). Internal consistency reliabilities are around .7 and correlation with verbally oriented intelligence tests is about .5. Correlation with the WCGTA is .48 and split half reliabilities range from .55 to .76 (Rane-Szostak & Robertson, 1996). Subscale scores should not be used on an individual basis due to the lack of evidence from factor analysis studies that they measure distinct skills. User norms are employed and are not representative of a well-defined group. A panel of twelve psychologists published in critical thinking evaluated the CCTT. Since it met the criteria for only 4 of the 15 validity and reliability standards from the 1974 Standards for Educational and Psychological Tests, it needed revision to

improve reliability and validity (Jacobs, 1995). Two reviewers, Hughes (1995) and Malcolm (1995) both cite the need for evidence of construct validity. Rane-Szostak and Robertson (1996) and Ennis (1993) note that this test may better serve as a teaching tool. This test focuses on evaluation rather than the productive aspect of critical thinking and does not test critical thinking dispositions (Rane-Szostak & Robertson, 1996).

Few nursing studies have used the CCTT. Allegretti and Frederick (1995) found significant differences in scores on the CCTT three months after a group of nursing students were taught the Toulmin, Rieke, and Janik Critical Thinking Model, in an interdisciplinary course on ethical reflection (Allegretti, 1995). Dungan (1985) found no significant difference between freshmen, second year ADN and senior level BSN students on the CCTT. Higher education level was related to better clinical judgment scores. The author raised questions about the strength of the CCTT as a means of measuring the quality of critical thinking in nursing (Pless & Clayton, 1993). Farley and Elmore (1992) looked at the relationship between critical thinking skills as measured by the CCTT and reading comprehension for underachieving college freshmen. No significant relationships were found between reading comprehension and critical thinking skills, but the planning experiments subscale was significantly related to reasoning in reading.

Due to its inferior rating to the WGCTA on the essential standards for educational and psychological tests, and reviewers' concerns about lack of construct validity, the CCTT has not developed strong support in research on nursing students. It can be used as suggested as a teaching tool instead.

### The Ennis Weir Critical Thinking Essay Test

The Ennis Weir Critical Thinking Essay Test, developed in 1985 for 7<sup>th</sup> grade – adult differs from the majority of other tools. It does not ask for bound responses, but requires constructive responses to evaluate written arguments found in an essay. Skills assessed include critiquing thinking, getting to the point, seeing reasons and assumptions, stating one's point, offering good reasons, seeing other possibilities, responding appropriately to and avoiding equivocation, irrelevance, circularity, reversal of if-then relationships, overgeneralization, credibility problems, use of emotive language to persuade, excessive skepticism, and straw-person fallacy (Adams, Whitlow, Stover, & Johnson, 1996; Ennis, 1993; Tompkins, 1989). Raters who have completed at least a college level course in logic or critical thinking must evaluate the students' responses (Jacobs, 1994; Rane-Szostak & Robertson, 1996). Interrater reliability estimates were .86 and .82 which is high for essay tests but only means that the raters ranked the students consistently, not that similar scores were given (Norris & Ennis, 1989). Content validity is used but construct validity has not been studied (Poteet, 1989). Most critical thinking instruments use conclusions to thinking processes rather than information about what the examinee is really thinking. Essay tests help provide this information. This test also differs in that it is criterion-referenced, not norm-referenced.

Due to limited validity information and lack of reliability, one reviewer suggests that this test may also be better used as a teaching tool (Tompkins, 1989). Because of these concerns and probably due to difficulty grading the essay questions, this tool has not been useful in nursing. No nursing studies were found that used this tool.

## The California Critical Thinking Skills Test & The California Critical Thinking Dispositions Inventory

Few critical thinking tools are based on research to the extent that the California Critical Thinking Skills Test (CCTST) and California Critical Thinking Dispositions Inventory (CCTDI) were. The CCTST, developed in 1990 by Peter Facione, uses the Delphi definition of critical thinking as its conceptual basis and addresses the cognitive aspect of critical thinking (Facione & Facione, 1994). It has been characterized as the best commercially prepared critical thinking tool in comparative analyses (P. A. Facione, 1991; Rane-Szostak & Robertson, 1996). The CCTST was designed for evaluation of programs, student assessment and research, and is intended for native English speakers at college and adult levels (P. A. Facione, 1991). Facione, Facione, & Sanchez (1994) note that the Delphi description of critical thinking attributes describes the characteristics of a nurse with ideal clinical judgment. Facione and Facione (1994) believe that the CCTST can serve nursing well. The CCTST and the CCTDI are the only standardized critical thinking tests specifically listed in the NLNAC interpretive guidelines for accreditation criteria.

The format covers situational, discipline neutral content with neutral and controversial general, familiar topics (P. A. Facione, 1991). It consists of 34 multiple-choice items in two forms. The questions are challenging, take 45 minutes, resemble a reading comprehension test, and can be used for post-test only or pretest-posttest designs (California Academic Press, 1996b, 1996c). There are three subscales in the 34 items, analysis, evaluation and inference, and two other subscales of deductive reasoning and inductive reasoning in 29 of the items (Jacobs, 1995). Scores for the subscales are

separate and should not be used for more than aggregate strengths and weaknesses (Jones & Brown, 1993).

Facione and Facione (1994) describe the development, validation, and pilot testing of the CCTST. Technical studies have been completed on construct, content, and concurrent validity, item discrimination and difficulty, and reliability (P. A. Facione, 1991). The CCTST was constructed from a bank of 200 items which were piloted, analyzed, reviewed and culled until a final test of 34 items was determined. Items were selected for the CCTST for their ability to cover the domain of the five cognitive critical thinking skills (Facione & Facione, 1994). The K-R 20 coefficients for all items are .69 for the pretest and .68 for the posttest, with a pre-post correlation of .70 (Adams, Whitlow, Stover, & Johnson, 1996). No reliabilities are given for the subscale scores (Jacobs, 1995). Increasing the length of the CCTST to increase reliability would affect its ability to be given during a typical class period. It would also likely cause increased mental fatigue due to the difficulty of the questions and result in decreased reliability estimates (Facione & Facione, 1994). The test does not differentiate unfairly between gender, ethnicity, level of critical thinking confidence, and major, but all genders and ethnic groups do not have equal results following critical thinking instruction (P. A. Facione, 1991).

Attempts were made to construct a parallel form item by item with the result of 28 changed items and 6 unchanged items. Equivalence is argued on conceptual grounds using within form data from two student groups. The alpha reliability is estimated at .71 (Jacobs, 1995). However, a study of 1,383 students at a large university found no significant difference on the unchanged items and a significant difference in means for



Form A and B, with form B being more difficult than form A (Jacobs, 1995). The test manual states that the two forms are statistically equivalent and the same for practical purposes (California Academic Press, 1996c). The differences between forms show how critical thinking is highly contextually sensitive (Jacobs, 1995). It is suggested that the total scores may be adequate for research purposes but decisions should not be made about individuals on the basis of scores from both forms since comparability is questionable (Jacobs, 1995).

Content validity of the CCTST is based on its relationship to the American Philosophical Association Delphi study research and the consensus of experts (Facione & Facione, 1994). Construct validity is supported in validation studies by small but significant gains in pretest-posttest scores in experimental groups after a critical thinking course was taken. Controls did not make the same gains (Facione & Facione, 1994). These gains were only between 0.04 and 1.45 in mean scores, but they were statistically significant for the large samples used (McMorris, 1995). This may explain why significant changes are not seen in many of the smaller scale studies. Construct validation is also supported by the high and significant correlation between the CCTST and the CCTDI reported in pilot and study samples (Facione & Facione, 1994). Additional research on construct validity is needed (Michael, 1995).

There is a question about the unique contribution of CCTST scores because of the high relationship between all subtest scores, total scores, SAT verbal scores, SAT math scores, Nelson-Denny Reading scores, and college GPA (Jacobs, 1994; California Academic Press, 1996c; Facione & Facione, 1994; McMorris, 1995).

The researchers raised a concern about students' low motivation for posttests. Students not in the validation study who were told that the test was their final improved their scores by a lot more. P. A. Facione (1991) concludes that due to the Delphi research finding and quantitative validation studies, core critical thinking skills can be taught, learned and objectively assessed. No evidence was found that critical thinking skill development is a natural outcome of baccalaureate education. P. A. Facione (1991) suggests that research checking the correlation between CCTST and other critical thinking assessment tools would be of interest. Studies need to control for selection, maturation, mortality, experimental effect, Hawthorne effect and other threats to internal and external validity (P. A. Facione, 1991).

Norms are based on the performance of students from a comprehensive urban state university (P. A. Facione, 1991). Facione suggests use of local norms because testing is controlled on site (Facione & Facione, 1994). Local norms may be created with at least 500 scores (P. A. Facione, 1991). McMorris (1995) advises test users to be cautious about interpreting results as this test does not have the history, reliability and variety of norms as does the WGCTA. Ennis (1993) lists its strengths as its sophistication, current theory base and alternative forms, with a limitation being the possible need for a specific critical thinking course for significant changes in scores to occur.

The CCTDI assesses dispositions toward critical thinking identified in the Delphi report as exemplifying the ideal critical thinker (California Academic Press, 1996a). The tool, which takes about 20 minutes, is designed for evaluation of groups and programs, research, and personnel and management training, and is targeted to high school, college,

and professional and graduate programs. It consists of 75 items with a six point Likert agree-disagree scale. Cronbach's alpha is reported at .9 overall with .72 to .80 for subscales (California Academic Press, 1996a; Facione, Facione, & Sanchez, 1994). Introduced in 1992, the CCTDI was developed to measure attitudes, beliefs and inclinations to use critical thinking (California Academic Press, 1996c). The scales are discipline neutral and are aimed at the baccalaureate level (Facione, Facione, & Sanchez, 1994). Concurrent validity is supported by significant correlation with established psychological scales targeting close constructs. In addition, two studies show highly significant correlation between scores on the CCTST and CCTDI (Facione & Facione, 1994; Facione, Facione, & Sanchez, 1994). Factor analysis was used in constructing the CCTDI. From 150 items with the strongest face validity, factor analysis determined 75 items which loaded the highest on seven factors for a pilot group (Facione, Facione, & Sanchez, 1994).

Studies are beginning to be reported in the nursing literature using the CCTST and CCTDI due to their relatively recent development. Table 2 summarizes results of nursing studies using the CCTST. Longitudinal studies using the CCTST don't appear in the literature until 1997. Data collected from 1992 to 1997 provided the largest aggregation of critical thinking skills and disposition data to date for fifty nursing programs using CCTST results (Facione, 1997). A significant increase in CCTST scores was noted from entry to exit.

Studies conducted by other than Facione and Facione have produced mixed results. Leppa (1997) used the CCTST to measure critical thinking skills for a BSN completion program in a pretest-posttest design study over a ten-month period. Results

Table 2 – Summary of Nursing Studies on Critical Thinking Using the California Critical Thinking Skills Test

Researcher	Type of Study	Sample	Major Findings about Critical Thinking	Variable showing relationship to critical thinking	Variables showing no relationship to critical thinking
Facione, 1997	Longitudinal	625 BSN	Significant increase from entry to exit		
Leppa, 1997	Longitudinal	70 RN-BSN	Scores decreased with significant decrease for inference scores		
Pepa, Brown, & Alverson, 1997	Longitudinal	45 Traditional BSN 43 Accelerated BSN	Accelerated group significantly higher than traditional at entry, but no difference at exit; Significant increase for traditional from entry to exit		
Thompson & Rebeschi, 1998	Longitudinal	38 BSN	Significant difference in total scores from entry to exit	Ethnicity	Age, Gender, GPA
Hopson, 1993	Cross-sectional	51 BSN 92 ADN	BSN scored significantly higher than ADN	Age (+) for ADN, Age (-) for BSN	
Lacey, 1996	Cross-sectional	93 BSN & ADN 30 Faculty	Significant difference between students and faculty but not between ADN & BSN		

Researcher	Type of Study	Sample	Major Findings about Critical Thinking	Variable showing relationship to critical thinking	Variables showing no relationship to critical thinking
Obenauf, 1996	Cross-sectional	335 BSN (4 levels from freshmen to senior)	No significant difference between the 4 levels	Academic demographic variables	
Colucciello, 1997	Cross-sectional	94 Sophomore II BSN 65 Junior I BSN 64 Junior II BSN 59 Senior I BSN 46 Senior II BSN	Junior I, Senior I & II, significantly higher than Sophomore II; Junior I had highest means		
Facione, 1997	Cross-sectional	311 Freshmen pre-BSN 1485 Sophomore BSN 1618 Junior BSN 2611 Senior BSN 633 MSN	Modest increases in cross-sectional data	Nelson-Denny Comprehension, GRE, ACT, SAT verbal, SAT math, GPA (5 kinds), Adopting a CT focus for curriculum, faculty discussing meaning of CT, faculty engaging in planning for CT, having a designated CT course, having analyzed CT assessment data, student faculty ratio (negatively)	RN status, Age, Sex, Nelson-Denny Vocabulary, having made curricular changes, designating CT learning outcomes,
Sorensen Bowles, 1997	Correlational	65 Senior BSN	Significant positive relationship between critical thinking and clinical judgment	GPA	

showed a slight decrease in total scores, with a significant decrease in the inference subscale. The mean pre-test scores of 18.85 were higher than the norm of 15.89. Researchers were concerned about the low reliability of subscales ranging from .21 to .51 making it difficult to tell what the scores actually meant. This school dropped the CCTST as a measure of critical thinking due to its unsuitability for their program. Better success was found using the CCTDI as a pretest-posttest. Results on the CCTDI were all in a positive direction. All subscales except for inquisitiveness had statistically significant increases. The program decided to use graduate portfolios to demonstrate evidence of critical thinking in course work. The Holistic Critical Thinking Scoring Rubric helped the faculty evaluate the portfolios and improve interrater reliability (Leppa, 1997).

In a comparative study of curricula types, Pepa, Brown, and Alverson (1997) studied critical thinking ability of students in traditional and accelerated BSN programs. Accelerated students had at least 44 credits and completed the nursing program in less than 22 months, which is faster than usual. There was a significant difference between the two groups at the beginning, but not at the end. This supports the assumption that students with more general education are better able to think critically. Because of the differences at the beginning, there were significant pre and post differences for traditional but not accelerated students, meaning that the traditional students caught up by the end of the program (Pepa, Brown, & Alverson, 1997).

Thompson and Rebesch (1998) used the CCTST and the CCTDI for BSN students on entry and exit. They found significant differences in total score on the CCTST and the total score and truth seeking and analyticity subscales on the CCTDI. No

difference was found for age, gender, or GPA, but a difference was found in ethnicity on the CCTST scores.

In the first correlational study to use CCTST, Hopson (1993) found that BSN students scored significantly higher than ADN students on the CCTST. Age was positively correlated for ADN students but negatively correlated for BSN students.

Lacey (1996) found a statistically significant difference in critical thinking skills between students and faculty for the CCTST but not between ADN and BSN students. No significant relationship was found between critical thinking ability and disposition. For critical thinking dispositions faculty scores were significantly higher than ADN students but not BSN students. Obenauf (1996) found significant differences between four class levels of BSN programs (freshmen through seniors) and critical thinking dispositions using the CCTDI but no significant differences between the four class levels and critical thinking skills using the CCTST.

In a cross-sectional, descriptive, comparative study with a non-randomized sample, Colucciello (1997) found statistically significant differences in critical thinking skills among five levels of BSN students on the CCTST and CCTDI. The junior 2 level had the highest mean score, and the overall scores of junior and senior I and II levels differed from the sophomore II level. All levels had weaknesses in truth seeking. Results showed a significant positive relationship between critical thinking skills and dispositions (Colucciello, 1997).

One correlational study using CCTST found a weak, but significant positive relationship ( $r = .21$ ) between critical thinking as measured by the CCTST and clinical

judgment using the Clinical Decision-Making in Nursing Scale (Sorensen Bowles, 1997).

A significant relationship was found between GPA and critical thinking scores.

In the Facione (1997) cross-sectional nursing metastudy, significant relationships were reported between scores on the CCTST and the CCTDI, and a variety of student descriptors such as SAT scores, GRE scores, Nelson-Denny Comprehension, GPA, adopting a critical thinking focus for the curriculum, faculty discussing the meaning of critical thinking, engaging in planning for critical thinking, and having a designated critical thinking outcome. No relationships were found between CCTST scores and RN status, age, sex, Nelson-Denny vocabulary, and having made curricular changes. There were modest increases found in cross-sectional data from freshmen to senior BSN students (Facione, 1997).

The CCTST is the most recently published of the general critical thinking tests. Concern has been raised over the relatively low K-R 20 reliability of .7. This is difficult to compare to the WGCTA because of the different methods of measuring reliability. The WGCTA has .69 to .85 reliability via the split half method. The K-R 20 method is expected to be lower with heterogeneous tools as each item is not expected to have high correlation with all the other items. Subscale reliabilities of both the WGCTA and the CCTST are low. Frisby (1991) noted that critical thinking tests constructed on a multidimensional framework may result in internal consistency reliability estimates and validity coefficients that are quite low. P. A. Facione (1990) notes that empirical research on how subscales correlate with each other and dispositions need to be done. He advises caution when interpreting critical thinking scores. Nursing schools are more likely to use the tests for aggregate data, rather than individual student scores, making the absolute



reliabilities not as crucial an issue. The CCTST has considerable content validity, but it needs further research to provide additional evidence of construct validity. Researchers other than the author of the test and his family should do this.

### Other Critical Thinking Assessment Tools

There are several other tools available for assessing critical thinking. The Framework for Externalizing Critical Thinking is a list of criteria that can be used to evaluate scientific presentations, clinical conferences, or class presentations to assess the quality of thinking leading to the project (Facione & Facione, 1996a). The Triarchic Test of Intellectual Skills, based on Sternberg's Triarchic theory of intelligence is a psychologically derived critical thinking test for high school and college levels. No normative, reliability or validity information is reported (Sternberg, 1986). Sternberg (1986) raises the concern that what this test measures is not clearly separate from intelligence.

The New Jersey Test of Reasoning Skills, developed by Virginia Shipman of ETS for the 4<sup>th</sup> grade through college level has 50 items that test for syllogistic reasoning, contradictions, caused relationships, assumption identification, induction, good reasons and other topics (Carpenter & Doig, 1998; P. A. Facione, 1996; Norris, 1987; Paul & Nosich, 1992). This philosophically derived test assesses 22 different skill areas with heavy emphasis on deduction (Norris & Ennis, 1989). Reliabilities are in the .6 to .8 level and correlation with reading comprehension is at the .8 level. This raises the concern that the variable measured is not clearly differentiated from verbal skills (Sternberg, 1986).

Some tools incorporate critical thinking as part of a variety of measurements. The Education Testing Service Academic Profile Tests measure critical thinking, reading, writing, and math within the context of humanities, social science, and natural sciences (P. A. Facione, 1996). The American College Testing Collegiate Assessment of Academic Proficiency (ACT CAAP) assesses six areas including critical thinking to evaluate general education programs (Yarbrough, 1992). This is the most frequently used standardized measure of general college outcomes (Rane-Szostak & Robertson, 1996). K-R 20 reliabilities range from .76 to .95, but limited validity information is given. Various norms are available. Concurrent correlation between the CAAP and GPA is modest. The critical thinking section covers analyzing, evaluating and extending arguments (Owen, 1998; J. Smith, 1998). In a study estimating the relative importance of three variables on the critical thinking abilities of students in the first year of college using the CAAP, results showed that students' level of involvement in and out of class has important effects on cognitive development including critical thinking, as it explained 52% of the variance in critical thinking scores (Terenzini, Springer, Pascarella, & Nora, 1995).

Several studies report use of instruments other than the commercial instruments frequently mentioned. Dorothy del Bueno found no significant differences between ADN, BSN, and diploma graduates' critical thinking abilities using a series of video simulations. She found that 62% were not performing at entry level (del Bueno, 1994). Keeley, Browne, and Kretuzer (1982) compared freshmen and seniors on general and specific open-ended questions and broad essay tests and found that both had significant deficiencies. Seniors scored slightly higher and differences were greatest when

instructions were more open-ended. A study at a Midwest university measured critical thinking ability of college seniors by an exercise evaluating critical thinking. Most were able to recognize errors in statistical reasoning, but not ambiguity in the use of language, questionable assumptions, and value proficiency. This showed that a traditional curriculum does not guarantee the internalization of critical thinking ability (Weddle, n.d. 1998, March 5).

Concern about the limitations of structured response formats for measuring critical thinking led Gunn (1993) to create a constructed response critical thinking assessment instrument, the Wasatch Test of Critical Thinking. The interrater reliability level was .74. Results showed the test was only marginally correlated with multiple-choice critical thinking tests at .36. A significant effect was found between critical thinking and academic class. The instrument was based on a definition of critical thinking that included evaluating products of thought based on judgment of the evidence and assessment of reasons, and the process of developing logical arguments.

Other tests have been designed primarily for lower-age students such as the Cornell Critical Thinking Test – Level X, the Test of Inquiry Skills, Judgment: Deductive Logic and Assumption Recognition, the Ross Test of Higher Cognitive Processes, the Basic Skills Assessment, Test of Cognitive Skills, Test of Problem Solving, Corrective Reading Mastery Test, Deductive Reasoning Test, the PSI Basic Skills Test for business and industry, Ball's Aptitude battery, and the Whimby Analytical Skills Inventory. Several tests evaluate only one aspect of critical thinking such as the Cornell Class Reasoning Test, the Cornell Conditioning Reasoning tests, the Logical Reasoning Test,

the Test on Appraising Observations and the Ennis Weir Argument Class Test (Ennis, 1993; Norris, 1987; Norris & Ennis, 1989; Paul & Nosich, 1992).

Besides Dorothy del Bueno's use of simulated videos, none of these tests have been used in nursing. The majority of these tools were developed for elementary and secondary level and have limited data on validity. None seem promising for nursing research.

### Discipline Specific Critical Thinking Assessment Tools

Some experts have called for discipline specific critical thinking tests (Perciful & Nester, 1996). Ford notes that the paradigm shift away from equating critical thinking with problem solving necessitates the development of a domain specific instrument to measure critical thinking in nursing (Vaughan-Wrobel, O'Sullivan, & Smith, 1997). In response to a need for a discipline specific critical thinking assessment tool for nursing, the Arnett Development Corporation developed the Critical Thinking Outcome Evaluation (CTOE) in the spring of 1998. This tool uses written response format instead of multiple-choice questions, and a partial credit model for evaluation. Each of the 15 items may have between eight and ten correct answers. Free entry eliminates guessing, and maximizes the ability to assess a student's true critical thinking ability (Arnett, 1998a). The test is designed to assess an individual's ability to use critical thinking, clinical decision making and clinical judgment in nursing situations. Each question uses content every graduate should know, and is categorized according to critical thinking competencies, categories of client needs, Bloom's Taxonomy, clinical area, systems, and ability level using item response theory (Arnett, 1998b). The tool is based on the Delphi

definition of critical thinking and measures five of the core critical thinking competencies identified by the Delphi study, analysis, interpretation, evaluation, inference, and explanation (Arnett Development Corporation, n.d. 1998, March 7).

The questions for the test were validated at the end of 1997 with 642 graduating students from BSN, ADN, and diploma programs and items were revised for the final tool based on statistical analysis (Arnett, 1998b). On-site testing with over ten thousand students in four basic nursing programs (PN, Diploma, ADN & BSN) took place in 1998. Usable results were obtained from 8,937 graduates (2,610 BSN, 4,491 ADN, 437 Diploma, and 1,399 PN). The following means were achieved: BSN – 79.3838; ADN – 79.06699; Diploma – 79.5023; and PN – 72.82867. Standard deviations ranged from 17.15946 to 18.33557. Scores ranged from minimums of 14-22 to maximums of 126-140. Split-half reliability for the test was .88. Experts contributed to the scoring key and content validity by contributing to a list of all possible correct answers. Answers were prioritized, and assigned point values. The answers were compared to student responses by the computer. A live rater evaluated any discrepancies between the computer and the student. Inter rater reliability was .95 because of a very small number of raters with high congruence. A correlation of .18 was found between the Arnett Computer Adaptive Test (CAT) which predicts success on the licensing exam and the CTOE. This was not unexpected and is explained because these tools test different things and have different formats. In 1999, schools will compare results of the next graduating class to those obtained the previous year (Arnett, 1998b). Additional studies need to be done to develop construct validity.

InterEd recently began developing profession specific critical thinking tests using Michael Scriven's multiple ranking methodology rather than exact answers. This method assesses two dimensions simultaneously, producing four to six data points for each one data point obtained in a multiple choice format (Tucker, 1996). There are advantages to assessing critical thinking in the context in which it occurs. This test focuses on a three element model of agency (dispositions), adequacy (intellectual capabilities), and context (Tucker, 1996a). The Critical Thinking Assessment for Nursing Education, the CTn, was the first of the discipline specific tests developed in August of 1996 ([InterEd@InterEd.com](mailto:InterEd@InterEd.com), n.d. 1998a, March 9). It measures critical thinking skills and attitudes that relate to nursing centered critical thinking processes. The case oriented multiple ranking questions and scaled attitudinal questions and response sets take about 40 – 50 minutes to complete. Scale alpha coefficients range from .68 to .92 and test reliability coefficients are above .75. The author reports high convergent validity where similar constructs are measured on existing tests and high discriminate validity when constructs are unique to the CTn ([InterEd@InterEd.com](mailto:InterEd@InterEd.com), n.d. 1998b, March 9). InterEd reports that the test is used in 32 NLN accredited programs in 21 states ([InterEd@InterEd.com](mailto:InterEd@InterEd.com), n.d. 1998c, March 9). Its use has not been mentioned in any of the current nursing critical thinking studies.

The International Center for the Assessment of Higher Order Thinking will devise an essay test for a program that directly tests for critical thinking skills in general or in a specific subject area. The program's faculty can learn how to grade the essays and develop student profiles based on written guidelines, the center's staff can grade them for a fee, or Richard Paul can provide an inservice on grading for an honorarium and

expenses (Paul & Elder, 1996). No reliability or validity information is given for the tests.

Given the relative newness of the standardized discipline specific critical thinking tests, and the debate on critical thinking as a general process or process that requires knowledge of a subject area, research is needed on these tools and their appropriateness for measuring critical thinking in nursing. How they relate to the general critical thinking skills tests must also be measured.

#### Summary and Conclusions from the Literature Review

Most of the nursing research studies in the literature using standardized critical thinking tests utilized the WGCTA. Convenience and often small samples limit the generalizability of results of those studies. Weaknesses in research design include lack of control groups to separate out maturational and general college effects, lack of randomization to groups, and failure to control for extraneous variables such as age and experience (Howenstein, Bilodeau, Brogna, & Good, 1996).

All the available standardized critical thinking assessment tools, including the WGCTA, have produced inconsistent results. The CCTST, one of the newer assessment tools, has shown the same kind of inconsistency. The level of difficulty of the questions on the CCTST tends to result in low scores and a difference of one question can mean a great difference in percentile ranking, which is a definite weakness. Strengths include its strong research base and the consensus definition on which it is based. The inconsistent results may be due to flaws in research design so it will take additional research over time to see if the CCTST has value as an assessment tool.

Several of the non-nursing studies mentioned deficiencies found in students' critical thinking. This was avoided in most nursing studies by the use of norm-referenced tests but it needs to be considered. Knowing that group A performed better than group B does not mean as much if both groups are below the expected standard. Emphasis in most nursing studies was on measuring significant differences between entry and exit or showing a difference based on educational level, not whether an acceptable level of critical thinking skills was demonstrated. This is a weakness in the use of norm-referenced tools. If critical thinking is a mandated outcome for accountability, studies should address achievement against a standard.

Very few of the critical thinking studies in the nursing literature used ADN students. When ADN students were included, it was usually in cross-sectional studies comparing them to BSN students. Little is known about the actual level of critical thinking skills in ADN students and what factors are associated both positively and negatively with critical thinking scores for those students. Current norms for the CCTST are for university level only. Norms are also needed for the ADN level. With ADN programs facing inclusion of critical thinking outcome assessment for NLNAC accreditation, research and norms for ADN students are all the more important.

A new generation of critical thinking assessment tools, discipline specific tools, is becoming available. Some are so new that no research has been reported yet. Both the newer and the current critical thinking tools need additional research to add to evidence of reliability and validity. From the critical thinking literature, it is clear that better evidence of construct validity is needed for most critical thinking assessment tools. Reliability of subscales is another area in which most critical thinking tools need to be



improved. Correlation between a new test and a previously available test is frequently cited as evidence of validity (Anastasi, 1982). When studies have compared the correlation between two critical thinking tests, concurrent validity between them has often been lower than expected. This has been explained because the tests had different scopes and were based on different theoretical constructs. No research was found that compared a discipline specific test and a general critical thinking skills test, especially ones that are based on the same definition of critical thinking.

This study addresses some of the deficiencies found in critical thinking research by studying critical thinking in associate degree nursing students and evaluating, in a concurrent validity study, the correlation between a discipline specific test and a general critical thinking skills test that are both based on the same definition of critical thinking. This study attempts to add a small piece of information to the growing body of knowledge about critical thinking. The more that is learned about critical thinking, the more complex the topic gets. There has been some progress in the development of consensus regarding what critical thinking is, and what skills it includes. What has been more elusive is how to measure it. With the divergence to discipline specific assessment tools, the lack of consistent results with any of the current tools, and the high correlation with many academic variables, consensus on measurement of critical thinking seems an elusive goal. However the requirement of critical thinking as a program outcome for accreditation purposes compels further study and a continued search for the best way to assess this complex construct.

## CHAPTER III

### METHOD AND PROCEDURES

#### Research Design

This study is a concurrent validity study between a general critical thinking skills test, the California Critical Thinking Skills Test (CCTST), and a discipline specific critical thinking test, the Arnett Critical Thinking Outcome Evaluation (CTOE).

#### Instruments

##### The California Critical Thinking Skills Test

The CCTST, developed in 1990 by Peter Facione after more than two decades of research, and published by the California Academic Press, uses the Delphi definition of critical thinking as its conceptual basis and addresses the cognitive aspect of critical thinking (Facione & Facione, 1994). It was designed for evaluation of programs, student assessment and research, and is intended for native English speakers at college and adult levels (P. A. Facione, 1991). The current costs for the CCTST range from \$45 for 25 copies to \$250 for 200 copies. The format of the test covers situational, discipline neutral content with general, familiar topics (P. A. Facione, 1991). It consists of 34 multiple-choice items and is available in two forms, A and B. The questions are challenging, take 45 minutes, resemble a reading comprehension test, and can be used for post-test only or pretest-posttest designs (California Academic Press, 1996b, 1996c). There are three subscales in the 34 items, analysis, evaluation and inference, and two other subscales of

deductive reasoning and inductive reasoning in 29 of the items (Jacobs, 1995). The three main subscales reflect the Delphi definition of critical thinking. The analysis subscale includes analysis and interpretation cognitive skills. The evaluation subscale includes evaluation and explanation cognitive skills and the inference subscale reflects only the inference cognitive skill. Scores for the subscales are separate and should not be used for more than aggregate strengths and weaknesses (Jones & Brown, 1993). The results of the test include a total score and a score for the five subscales. The results are given in raw scores. Table 3 gives the maximum scores for the total and each subscale of the CCTST.

Table 3

Maximum Scores for the California Critical Thinking Skills Test

Total Score	Analysis	Evaluation	Inference	Deduction	Induction
34	9	14	11	16	13

Facione and Facione (1994) described the development, validation, and pilot testing of the CCTST. Technical studies have been completed on construct, content, and concurrent validity, item discrimination and difficulty, and reliability (P. A. Facione, 1991). The CCTST was constructed from a bank of 200 items which were piloted, analyzed, reviewed, and culled until a final test of 34 items was determined. Items were chosen for their ability to cover the domain of the five cognitive critical thinking skills (Facione & Facione, 1994). A pretest/posttest, case/control study produced evidence for the validity and reliability of the CCTST, assessed instrumentation effects, and measured changes after one course in critical thinking. The K-R 20 coefficients for all items were

.69 for the pretest and .68 for the posttest, with a pre-post correlation of .70 (Adams, Whitlow, Stover, & Johnson, 1996). No reliabilities were given for the subscale scores (Jacobs, 1995). Facione and Facione (1994) believe that increasing the length of the CCTST to increase reliability would affect its ability to be given during a typical class period and decrease its usability. It would also likely cause increased mental fatigue due to the difficulty of the questions and result in decreased reliability estimates. The test does not differentiate unfairly between gender, ethnicity, level of critical thinking confidence, and major, but all genders and ethnic groups do not have equal results following critical thinking instruction (P. A. Facione, 1991).

Originally only one form of the test was available. A parallel form was constructed item by item with 28 being changed and 6 unchanged. Equivalence is argued on conceptual grounds using data from two student groups. The alpha reliability is estimated at .71 (Jacobs, 1995). The test manual states that the two forms are statistically equivalent and the same for practical purposes (California Academic Press (1996c). However, a study of 1,383 students at a large university found no significant difference on the unchanged items and a significant difference in means for Form A and B. Form B was found to be more difficult than form A (Jacobs, 1995). The differences between forms show how critical thinking is highly contextually sensitive (Jacobs, 1995). Jacobs (1995) suggests that the total scores may be adequate for research purposes but decisions should not be made about individuals on the basis of scores from both forms since comparability is questionable.

Content validity of the CCTST is based on its relationship to the American Philosophical Association (APA) Delphi study research and the consensus of experts

(Facione & Facione, 1994). Construct validity is supported by small but significant gains in pretest-posttest scores in validation studies after a critical thinking course was taken. Controls did not make the same gains (Facione & Facione, 1994). These gains were only between 0.04 and 1.45 in mean scores, but they were statistically significant for the large samples used (McMorris, 1995). Construct validity is also supported by the high and significant correlation between the CCTST and the California Critical Thinking Disposition Inventory reported in pilot and study samples (Facione & Facione, 1994). Additional research on construct validity is needed (Michael, 1995). There is a question about the unique contribution of CCTST scores because of the high relationship between all subtest scores, total scores, SAT verbal scores, SAT math scores, Nelson-Denny Reading scores, and college GPA (Jacobs, 1994; California Academic Press, 1996c; Facione & Facione, 1994; McMorris, 1995).

Norms are based on the scores of students from a comprehensive urban state university (P. A. Facione, 1991). Facione and Facione (1994) suggest use of local norms because testing is controlled on site. Local norms may be created with at least 500 scores (P. A. Facione, 1991). McMorris (1995) advises test users to be cautious about interpreting the results as this test does not have the history, reliability and variety of norms as does the Watson Glaser Critical Thinking Appraisal. Ennis (1993) lists as its strengths sophistication, current theory base and alternative forms. A limitation is the possible need for a specific critical thinking course for significant changes in scores to occur.

The APA Delphi Report states that the critical thinking cognitive skills act in interdependent and interconnected ways. Facione and Facione (1994) note that a factor

analysis performed to differentiate the different skills into factors will likely fail because the skills have to be used interactively to answer the items. The authors suggest that the subscale scores should not be used for more than gross indicators of possible critical thinking strengths and weaknesses.

#### The Arnett Critical Thinking Outcome Evaluation Tool

The Arnett Development Corporation developed the Arnett CTOE in the spring of 1998. The current cost of the instrument is \$10 per student which includes scoring. This free entry, written response test was developed using the latest testing strategies. The free response format eliminates guessing, and maximizes the ability to assess a student's true critical thinking ability (Arnett, 1998a). Each of the 15 items may have between eight and ten correct answers. Up to ten points can be awarded for each item, for a total of 150 points. Experts contributed to the scoring key and content validity by developing a list of all possible correct answers. Answers were prioritized and assigned point values. The answers were compared to student responses by the computer. A live rater evaluated any discrepancies between the computer and the student. Inter-rater reliability was .95 because of a very small number of raters with high congruence.

The test is designed to assess an individual's ability to use critical thinking, clinical decision making and clinical judgment in nursing situations. The tool is also based on the Delphi definition of critical thinking and measures five core critical thinking competencies identified by the Delphi study, analysis, interpretation, evaluation, inference, and explanation (Arnett Development Corporation, n.d. 1998, March 7).

The questions for the test were validated at the end of 1997 with 642 graduating students from BSN, ADN, and diploma programs and items were revised for the final tool based on statistical analysis (Arnett, 1998b). On-site testing with over ten thousand students in four basic nursing programs (PN, Diploma, ADN & BSN) took place in 1998. Usable results were obtained from 8937 graduates (2610 BSN, 4491 ADN, 437 Diploma, and 1399 PN). The following means were achieved: BSN – 79.3838; ADN – 79.06699; Diploma – 79.5023; and PN – 72.82867. Standard deviations ranged from 17.15946 to 18.33557. Scores ranged from minimums of 14-22 to maximums of 126-140. Split-half reliability for the test was .88. A correlation of .18 was found between the Arnett Computer Adaptive Test (CAT) which predicts success on the licensing exam and the CTOE. This was not unexpected and is explained because these tools test different things and have different formats. In 1999, schools will compare results of the next graduating class to those obtained the previous year (Arnett, 1998b). Additional studies need to be done to develop construct validity.

### Population

The population for this study was associate degree nursing students in their last semester of nursing school in National League for Nursing Accrediting Commission (NLNAC) accredited programs in Texas. Associate degree nursing students in Texas complete a two-year curriculum leading to an associate degree in nursing. Graduates are recommended to the Board of Nurse Examiners for the State of Texas to take the National Council Licensure Examination for Registered Nurses (NCLEX-RN) to become registered nurses. All ADN programs in Texas are required to incorporate the same

competencies, the Nursing Education Advisory Committee (NEAC) competencies, into their curricula. Thirty-seven of the ADN programs in Texas are currently NLNAC accredited.

### Sampling Method

Due to the cost of the Arnett CTOE tool, a convenience sample of NLNAC accredited schools of nursing using or planning to use the Arnett CTOE in the spring of 1999 and willing to participate in the study was solicited. A survey of interest was sent to the thirty-seven NLNAC accredited ADN programs in Texas. Twenty-two surveys were returned with nine schools using or planning to use the Arnett CTOE tool and willing to participate, and an additional nine schools interested in participating but not planning to use the Arnett CTOE Tool. A convenience sample of students in their last semester of nursing school at the nine schools planning to use the Arnett CTOE was obtained. Students willing to complete the demographic form and allow their results on the critical thinking tests to be used in the study were included.

### Procedure for Data Collection

The following procedure for collection of data was followed:

1. After appropriate approvals were obtained, the researcher made arrangements with a contact person at each participating school, finding out how many students were interested in participating in the study.
2. The researcher purchased the CCTST booklets and contacted the Arnett Development Corporation regarding arrangements for obtaining the Arnett CTOE tools. Dr.



- Penelope Arnett of the Arnett Development Corporation volunteered to provide the CTOE booklets and scoring to the participating students free of charge.
3. The investigator mailed the demographic surveys with cover letters and consent forms explaining the study and requesting consent, the Arnett CTOE tools and the CCTST booklets and scantrons to the contact person at each school. A copy of the demographic survey, cover letter, and consent form are included in the appendices.
  4. The contact person administered the tests and demographic surveys to the students in the last semester of their ADN program during the spring of 1999. The schools were asked not to administer either of the tests in the last week of the semester. N. C. Facione (1997) found that there was no significant difference in scores on the CCTST given in the first half of the semester and the last half of the semester. The mean scores of tests taken in the last week of the semester were decreased.
  5. The contact person was responsible for assigning each student a code number for tracking purposes and making sure the CCTST scantron, the Arnett CTOE booklet and the demographic survey had the same code numbers.
  6. The contact person returned the demographic surveys, the CCTST booklets and scantrons to the researcher by the end of the spring 1999 semester. The researcher graded the CCTST scantrons using the PAR Score system and sent the results of the CCTST to the contact person at each of the participating schools by the end of the spring semester.
  7. The contact person returned the Arnett CTOE booklets to the Arnett Development Corporation by the end of the spring 1999 semester. The Arnett Development

- Corporation graded the tools and sent the results to the participating schools during the summer of 1999 and the raw data to the researcher by the fall of 1999.
8. The researcher performed the statistical analysis of the results.
  9. The researcher sent copies of the results of the analysis to the participating schools at the conclusion of the study.

### Data Analysis

The collected data were analyzed in the following manner:

Pearson product-moment coefficients ( $r$ ) between scores and subscale scores on the CCTST and the Arnett CTOE were calculated to test each section of hypothesis one.

Hypothesis two was tested by the following computations:

Hypothesis 2a was tested by calculating the Pearson product-moment coefficient ( $r$ ) between the age of the subjects and the scores on the CCTST and the Arnett CTOE.

Hypothesis 2b was tested by calculating the Pearson product-moment correlation coefficient ( $r$ ) between the scores on the CCTST and the Arnett CTOE and the status of completion of a previous degree. The Pearson correlation procedure is appropriate for variables on an interval scale. The CCTST and CTOE scores are on an interval scale. Degree status is a nominal variable. The point-biserial correlation analysis is appropriate for an interval variable and a dichotomous nominal variable. Degree status, as measured in the survey, had four levels and was not a dichotomous variable. For the statistical analysis, this variable was reduced to a dichotomous variable. The point-biserial correlation coefficient is a special case of the Pearson  $r$  and can be calculated by the same formula (Hinkle, Wiersma, & Jurs, 1994).

Hypothesis 2c was tested by calculating the Pearson product-moment coefficient ( $r$ ) between the GPA of the subjects and the scores on the CCTST and the Arnett CTOE.

Hypothesis 2d was tested by calculating the Pearson product-moment coefficient ( $r$ ) between the scores on the CCTST and the Arnett CTOE and the first generation in college status of the students in the sample. The point-biserial correlation coefficient ( $r_{pb}$ ), appropriate for a dichotomous nominal variable such as first generation in college status and an interval variable such as critical thinking scores, is calculated by the same formula as the Pearson product-moment correlation procedure.

Hypothesis 2e was tested by calculating the Pearson product-moment coefficient ( $r$ ) between the years of work experience of the subjects in a health care setting and the scores on the CCTST and the Arnett CTOE.

Hypothesis 2f was tested by calculating the Pearson product-moment correlation coefficient ( $r$ ) between the scores on the CCTST and the Arnett CTOE and the LVN status of the students. The point-biserial correlation coefficient ( $r_{pb}$ ), appropriate for a dichotomous nominal variable and an interval variable, is calculated by the same formula as the Pearson product-moment correlation procedure.

Hypothesis 2g was tested by calculating the Pearson product-moment correlation coefficient ( $r$ ) between the scores on the CCTST and the Arnett CTOE and the gender of the students. The point-biserial correlation coefficient ( $r_{pb}$ ), appropriate for a dichotomous nominal variable such as gender and an interval variable such as critical thinking scores is calculated by the same formula as the Pearson product-moment correlation procedure.

Hypothesis 2h was tested by calculating the Pearson product-moment correlation coefficient ( $r$ ) between the scores on the CCTST and the Arnett CTOE and the ethnicity of the students. Ethnicity is a nominal variable. The point-biserial correlation analysis is appropriate for an interval variable and a dichotomous nominal variable. Ethnicity, as measured in the survey, had six levels and was not a dichotomous variable. For the statistical analysis, this variable was reduced to a dichotomous variable. The point-biserial correlation coefficient is a special case of the Pearson  $r$  and can be calculated by the same formula (Hinkle, Wiersma, & Jurs, 1994).

Hypothesis 2i was tested by doing regression analyses with the three independent variables and the two dependent variables, the CCTST scores and the CTOE scores.

## CHAPTER IV

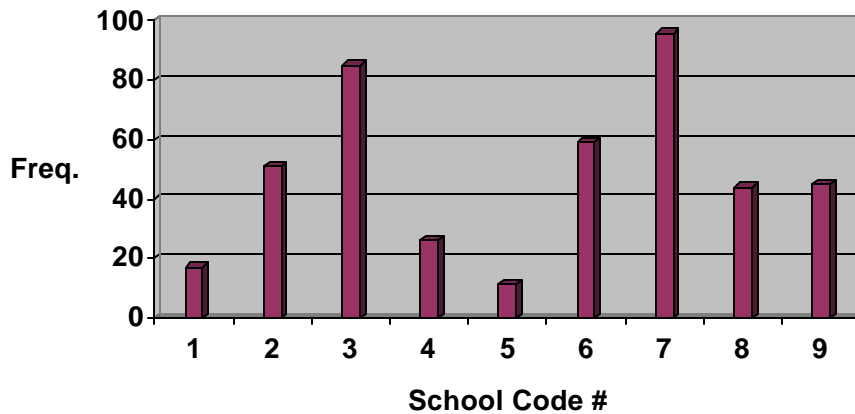
### RESULTS

#### Introduction

This study investigated the concurrent validity between two critical thinking tests, one a discipline specific test, the Arnett Critical Thinking Outcome Evaluation tool (CTOE), and the other a general critical thinking skills test, the California Critical Thinking Skills Test (CCTST). Four hundred and thirty-four associate degree nursing students enrolled in their last semester of nursing school completed the demographic questionnaire for this research study in the spring semester of 1999. This convenience sample consisted of student volunteers from nine community colleges in Texas. The participating colleges, Alvin Community College, Collin County Community College, Grayson College, McLennan Community College, North Central Texas College, San Antonio College, Tarrant County College, Trinity Valley Community College, and the University of Texas at Brownsville/Texas Southmost College, represent various regions of Texas. The size of the groups participating ranged from a low of 11 students to a high of 96. Figure 1 shows the distribution of students from the nine schools. The alphabetical listing of the schools above does not correlate with the numerical listing in the table.

Figure 1

Distribution of Participants by School



Of the 434 students, 4 students did not complete the CCTST and 11 did not complete the Arnett CTOE. Two students who took the CTOE put the same code number on their booklets, making their scores unusable. Scores from 417 students who took both tests were used to determine the correlation between the two tests and their subtests. On the demographic survey, data were collected to measure the independent variables of age, gender, GPA, health care employment experience, LVN status, additional degree status, and first generation in college status. These variables were analyzed to determine if they correlated with the scores on each of the critical thinking tests, the dependent variables. Correlation statistics were performed for various sets of variables. In addition, regression analysis was done to determine the significance of the predictor variables.

## Descriptive Statistics of the Sample

### Gender

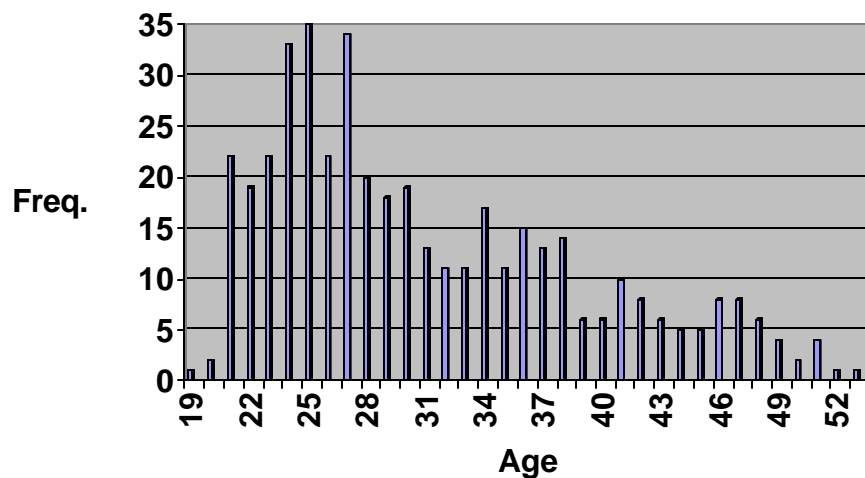
A majority of the sample, 372 or 85.7% were female and only 62 students or 14.3% were male. These findings were expected. Nursing remains a predominantly female profession. National Council Research and Statistical Data (1999) on newly licensed RNs in 1997 showed 87.9% female and 11.74% male.

### Age

The mean age of 432 students in the sample was 31, with a standard deviation of 7.97. Two students declined to answer the question of age. The median was 29 and the mode was 25. The ages ranged from the youngest at 19 to the eldest at 54. The graph in Figure 2 representing the ages of the students in the sample shows negatively skewed results with a majority of the students in their twenties. Almost 52% were between the ages of 21 and 29.

Figure 2

### Distribution of Ages of Students in Sample



The National Council Research and Statistical Data on age distribution by type of RN educational program showed that the mean age of 50,815 newly licensed RNs from ADN programs nationally in 1997 was 31.32 with a standard deviation of 8.13 (National Council of State Boards of Nursing, 1999a). A *t* test performed to compare the means of the ages of the national group and this study sample resulted in acceptance of a null hypothesis that the means were equal ( $t=.02314$ ;  $df=1$ ;  $CV = 12.7$  for significance at the .05 level). No descriptive data is available on students in their last semester of nursing school. The majority of the students in the sample would be newly licensed RNs in the summer of 1999, making a group of newly licensed RNs the closest group for comparison.

#### Ethnicity

The students in the sample represented a variety of ethnic backgrounds with the majority, 322 or 74.2%, being Caucasian. The rest of the sample consisted of 27 African-Americans (6.2%), 6 Asian or Pacific Islander (1.4%), 68 Hispanic (15.7%), 7 Native American (1.6%), and 3 other (.7%). One student declined to specify ethnicity. In the National Council data on newly licensed RNs in 1997, 80.42% were Caucasian, 6.23% were African American, 5.59% were Asian or Pacific Islander, 3.58% were Hispanic, .75% were Native American, and 1.07% were other (National Council of State Boards of Nursing, 1999b). The main differences between the groups were a lower percentage of Asian or Pacific Islander students in the study group and a higher percentage of Hispanic students. The higher Hispanic percentage was due to schools from predominantly Hispanic areas, San Antonio and Brownsville, participating in the study.

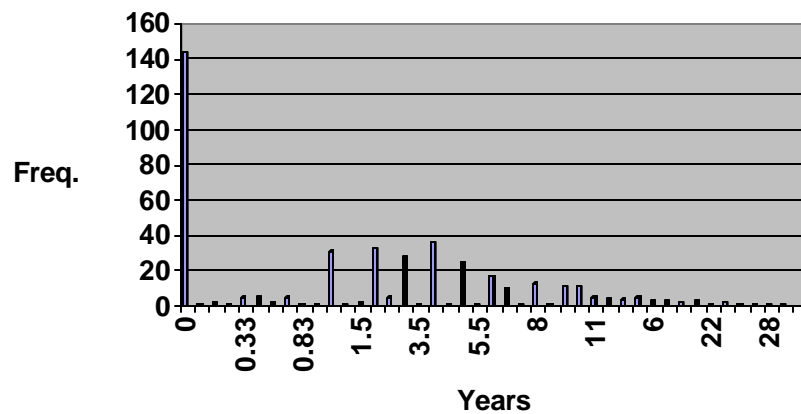


## Health Care Employment Experience

Health care employment experience ranged from a low of zero for a third of the sample (33.2%) to a high of 30 years for one student. Health care experience could have been in any of a variety of positions such as nurse aide, ward clerk, paramedic or EMT, LVN, or surgical technologist, etc. The study did not differentiate which position the student held. The mean experience was 3.7 years with a median of 2.00. Figure 3 shows the distribution of health care employment experience. Three students declined to answer this question.

Figure 3

### Distribution of Health Care Employment Experience in the Sample



## LVN Status & Transition Status

Ninety-seven or 22.4% of the students in the sample were current LVNs when they completed the demographic survey. The Board of Nurse Examiners for the State of Texas (1999) reported that of 3,083 ADN graduates from 9/1/97 – 8/31/98, 639 or 20.73% were LVNs prior to graduating. Of the 97 LVNs in this study, 77 or 80.2% went through an articulated LVN to ADN transition program. The other 19.8% either became a

LVN after completing a sufficient portion of the ADN program and therefore were newly licensed LVNs or they were LVNs before starting the ADN program from the beginning as a generic student. Future studies should better delineate the LVN status of students in the sample.

#### Prior Degrees

Sixty-eight students (15.7%) in the sample had completed one or more educational degrees prior to the ADN program. Thirty-three students or 7.6% completed a different associate degree, 32 or 7.4% completed a bachelor's degree and 3 students or .7% completed a master's degree. One student declined to answer this question. If a student completed more than one degree, only the highest degree was used in the data analysis.

#### First Generation in College Status

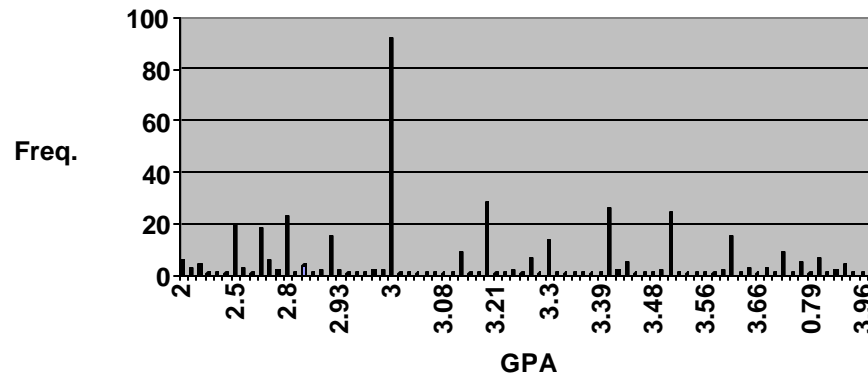
More than half of the students in the sample, 232 or 53.5% reported that their generation was the first to go to college.

#### GPA

The mean GPA of the sample of 424 students (10 declined to answer) was 3.1345. The median and the mode were 3.00. Eighty different GPAs were reported from a high of 4.0 to a low of 2.0. The accuracy of the GPAs obtained may be affected by the self-reported nature. Several students recalled exact GPAs but many estimated or rounded as evidenced by the high number of rounded numbers such as 2.50, 2.80, 3.00, 3.50, etc. Figure 4 represents the distribution of GPAs of the students in the sample.

Figure 4

Distribution of GPAs of the Students in the Sample



Although community colleges are open door by nature, most nursing programs have minimum GPA criteria for selection and several choose students by selecting students with the highest GPAs. An ANOVA analysis comparing the GPA means of the different schools showed that there was a significant difference in GPA between schools at the .01 level. Post hoc Sheffe test results showed multiple differences between schools. The test of homogeneity of variances (Levene statistic = 1.350; df1 = 8; df2 = 415; sign. = .217) showed no significant differences in the variance distribution between schools. Table 4 shows the results of the ANOVA analysis comparing the means of the GPAs of students by school.

Table 4

ANOVA Results for GPA Means by School

		Sum of Squares	df	Mean Squares	F	Sig.
GPA	Between Groups	16.959	8	2.210	17.738	.000
	Within Groups	49.597	415	.120		
	Total	66.556	423			

## Critical Thinking Test Results

## California Critical Thinking Test

Four hundred and thirty students took the CCTST. The results were analyzed using the ParScore system. The scores ranged from a high of 28 to a low of 4, with all the scores in between occurring at least once. The maximum score was 34. The mean of the sample was 15.0023 with a standard deviation of 4.1225. The median and mode were 15.00. The K-R 20 reliability of the CCTST was .61. Each of the items had a positive point biserial demonstrating good discrimination. Twenty-eight of the 34 items had discriminations over .20. Figure 5 shows the distribution of total scores on the CCTST and Table 5 shows the comparison of the CCTST total scores, subtest scores, means, standard deviations and the relationship of the mean to the maximum possible scores.

Figure 5

Distribution of California Critical Thinking Test Scores

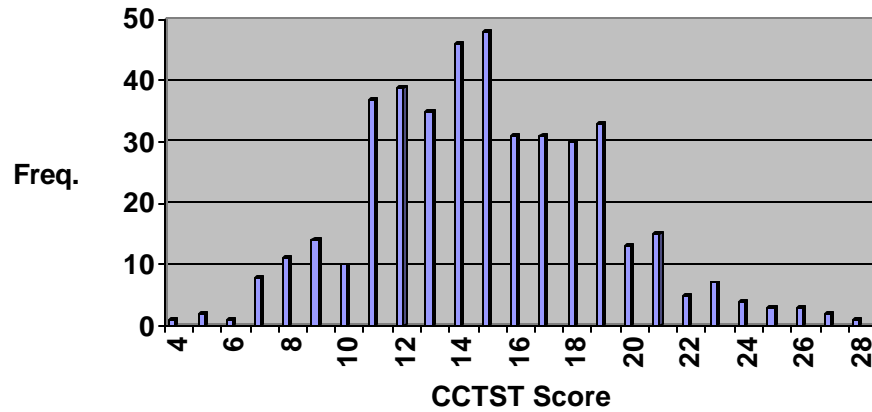


Table 5

Comparison of CCTST & Subtest Results

N = 430

	Total CCTST	Analysis	Evaluation	Inference	Deduction	Induction
Maximum Possible Score	34	9	14	11	16	13
M	15.0023	4.3698	4.9953	5.6372	7.2023	5.8093
SD	4.1225	1.5269	2.3937	1.7511	2.4954	2.1058
Rel. of mean to max. score	44%	48.55%	35.68%	51.25%	45.01%	44.69%

When the relationship of the mean to the maximum possible score was calculated, the students in the study sample scored best on the inference subscale and worst on the evaluation subscale. There are no CCTST norms for associate degree nursing students.

Comparing the results of this study with the University senior nursing students in the Facione (1997) metastudy, the total scores of this study's sample fell at about the 30<sup>th</sup> percentile, as did the evaluation and induction subscale scores. The analysis subscale scores were at about the 40<sup>th</sup> percentile and the remaining two subscale scores, inference and deduction were at the 50<sup>th</sup> percentile.

#### Arnett Critical Thinking Outcome Evaluation Tool

Four hundred and twenty-one students took the Arnett CTOE. The scores ranged from a high of 128 to a low of 28. The maximum possible score was 150 with a score of up to 10 points given for each of 15 questions. The mean score for the sample was 82.6888 with a standard deviation of 16.9417. A total of 43 different scores occurred. The distribution was slightly skewed to the right. Figure 6 shows the distribution of scores on the Arnett CTOE. The only norms to compare the results with are the outcomes from the 1998 research project (Arnett, 1998) in which ADN students scored a mean of 78.61 with a high of 90.7 and a low of 58.

The Arnett CTOE subtest results were reported as the number of questions missed in each of the critical thinking core competency categories. The higher the number of questions missed the poorer the score. If a student received a partial credit for a question it was not counted as missed. Table 6 shows the comparison of the total number of questions in each CTOE subcategory, the mean number of questions missed in each category, the standard deviations and the means from the 1998 research on ADN students for purposes of comparison.

Figure 6

Distribution of Arnett Critical Thinking Outcome Evaluation Scores

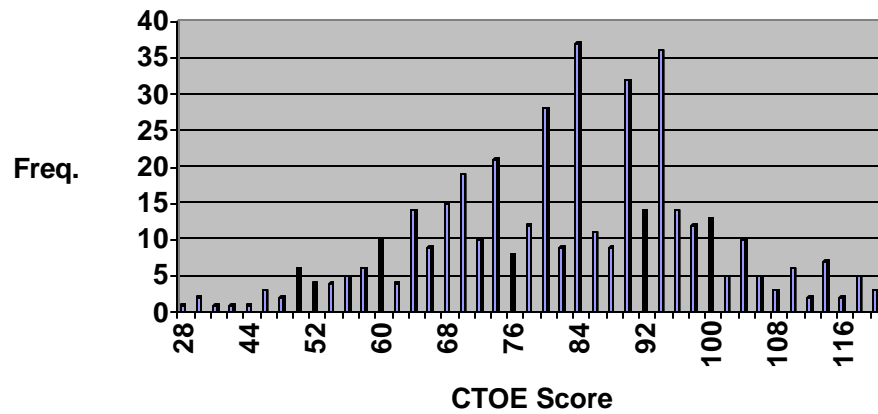


Table 6

Comparison of CTOE Subtest Results by Number of Questions Missed

N = 421

	Interpretation	Analysis	Evaluation	Inference	Explanation
Maximum Possible # Questions	3	3	3	4	2
Mean # of Questions Missed	.4418	1.1069	2.7387	1.0024	.7791
SD	.6091	.7241	.5505	.9011	.5905
ADN Mean # Questions Missed in 1998 Study	.618	1.36	2.74	1.04	.92

One student received partial credit on each of three questions in one subtest, scoring 4 out of 10 possible points for each question for a total of 12. Since the student had a partial score on each question, the student report noted that the student missed zero questions. Another student received full credit for two of the questions and no score for the third question for a total raw score of 20. This student had the higher raw score, but missed one question on the student report. Missing one question was noted to be a poorer score than missing zero questions. Due to this discrepancy, the raw data was obtained from the Arnett Development Corporation and included in the data analysis. Table 7 shows the mean raw scores and standard deviations for each of the subtests in the sample as well as the relationship of the means to the maximum possible score. No norms or comparable results were available for the raw scores.

Table 7

Comparison of CTOE Subtest Results by Raw Score

N = 421

	Total CTOE	Interpretation	Analysis	Evaluation	Inference	Explanation
Maximum Possible Score	150	30	30	30	40	20
M	82.7062	20.1330	18.9216	2.5891	28.7886	12.2090
SD	16.9253	6.5452	7.2418	5.4941	8.7899	5.9051
Rel. of Mean to Max. Possible Score	55.14%	67.1%	63.07%	8.63%	71.97%	61.05%



For the CTOE, the results of four of the subcategories were fairly consistent with students doing the best in the inference subcategory, the same as on the CCTST. However, the scores for the evaluation subcategory were very low. Evaluation was also the lowest subscale on the CCTST, but the difference was not as great. These results were similar to those from the 1998 research on the CTOE.

### Hypotheses Tested

#### Hypothesis One

- a. Hypothesis 1a stated there would be a positive correlation between the total scores of associate degree nursing students on the CCTST and the total scores on the Arnett CTOE. The Pearson correlation coefficient calculated on the total CCTST and CTOE scores for the 417 students who took both tests was .238, which was statistically significant at the .01 level (2-tailed). There was a very weak but positive correlation between the total CCTST and CTOE scores for the study group of ADN students. The null hypothesis was rejected and Hypothesis 1a was accepted.
- b. Hypothesis 1b stated that there would be a positive correlation between the analysis subscale scores of associate degree nursing students on the CCTST and the combined analysis and interpretation subscale scores on the Arnett CTOE. The Pearson correlation coefficient calculated for the analysis subscale scores of the CCTST and the combined analysis and interpretation subscores of the Arnett CTOE for 417 students was .048. The correlation was not significant at the .05 level. The null hypothesis was accepted and Hypothesis 1b was rejected.

- c. Hypothesis 1c stated that there would be a positive correlation between the evaluation subscale scores of associate degree nursing students on the CCTST and the combined evaluation and explanation subscale scores on the Arnett CTOE. The Pearson correlation coefficient calculated for the evaluation subscale scores on the CCTST and the combined evaluation and explanation subscale scores on the CTOE for 417 students was .210, which was statistically significant at the .01 level (2-tailed). This showed a very weak but positive correlation. The null hypothesis was rejected and Hypothesis 1c was accepted.
- d. Hypothesis 1d stated that there would be a positive correlation between the inference subscale scores of associate degree nursing students on the CCTST and the inference subscale scores on the Arnett CTOE. The Pearson correlation coefficient calculated between the inference subscale on the CCTST and the inference subscale on the CTOE was .099, which was statistically significant at the .05 level (2-tailed). There was a very weak but positive correlation. The null hypothesis was rejected and Hypothesis 1d was accepted.

## Hypothesis 2

Hypothesis 2 stated that intervening variables would be associated with critical thinking in associate degree nursing students in specified ways.

- a. Hypothesis 2a stated that age would positively correlate with critical thinking scores on a general critical thinking skills test and a discipline specific critical thinking test for associate degree nursing students. The Pearson correlation coefficient between age and CCTST total was .002, which was not significant at the .05 level. The Pearson correlation coefficient between age and CTOE total was -.007, which was not

significant at the .05 level. The null hypothesis was accepted and Hypothesis 2a was rejected for CCTST and CTOE.

- b. Hypothesis 2b stated that completion of a previous degree would positively correlate with critical thinking scores on a general critical thinking skills test and a discipline specific critical thinking test for associate degree nursing students. The Pearson correlation procedure is appropriate for variables on an interval scale. The CCTST and CTOE scores are on an interval scale. Degree status is a nominal variable. The point-biserial correlation analysis is appropriate for an interval variable and a dichotomous nominal variable. Degree status, as measured in the survey, had four levels and was not a dichotomous variable. For the statistical analysis, this variable was reduced to a dichotomous variable with one level meaning the student had no other degrees prior to the ADN program and the other level meaning the student had completed an associate, bachelors, or masters degree prior to the ADN program. The point-biserial correlation coefficient is a special case of the Pearson  $r$  and can be calculated by the same formula (Hinkle, Wiersma, & Jurs, 1994). The Pearson correlation coefficient calculated between completion of a previous degree and the CCTST total was .078, which was not significant at the .05 level. The Pearson correlation coefficient calculated between completion of a previous degree and the CTOE total was .011, which was also not significant at the .05 level. The null was accepted and Hypothesis 2b was rejected for CCTST and CTOE.
- c. Hypothesis 2c stated that GPA would positively correlate with critical thinking scores on a general critical thinking skills test and a discipline specific critical thinking test for associate degree nursing students. The Pearson correlation coefficient calculated

between GPA and the CCTST total was .254 which was significant at the .01 level.

The Pearson correlation coefficient calculated between GPA and the CTOE total was .192, which was also significant at the .01 level. The null hypothesis was rejected and Hypothesis 2c was accepted for both CCTST and CTOE.

- d. Hypothesis 2d stated that first generation in college status would negatively correlate with critical thinking scores on a general critical thinking skills test and a discipline specific critical thinking test for associate degree nursing students. Since the first generation in college status was a dichotomous variable and the point-biserial correlation coefficient has the same formula as the Pearson correlation coefficient, the Pearson correlation procedure was used to test this hypothesis. For dichotomous variables such as first generation in college status, LVN status, and gender the assignment of 1 and 2 is strictly arbitrary and the sign of the Pearson  $r$  depends on the meaning of this assignment (Hinkel, Wiersma, & Jurs, 1994). If the  $r$  is positive, the variable assigned the higher number is the one that is positively correlated with the interval variable. If the  $r$  is negative, the variable with the lower number is the one that is positively correlated with the interval variable. The Pearson  $r$  coefficient calculated between first generation of college status and the CCTST total was .135, which was significant at the .01 level. Since 1 = Yes and 2 = No for the first generation in college status, these results showed a slight negative correlation between CCTST scores and students who were in the first generation of their family to go to college. The null hypothesis was rejected and Hypothesis 2d was accepted for CCTST. The Pearson  $r$  coefficient calculated between first generation of college status and the CTOE total was .06, which was not significant at the .05 level. The null

- was accepted and Hypothesis 2d was rejected for CTOE, the discipline specific critical thinking skills test.
- e. Hypothesis 2e stated that the number of years of experience working in a health care setting would positively correlate with critical thinking scores on a discipline specific critical thinking skills test and not correlate with scores on a general critical thinking test for associate degree nursing students. The Pearson  $r$  coefficient calculated between number of years of employment in a health care setting and the CCTST total was  $-.183$ , which was significant at the  $.01$  level. This showed a slight negative correlation. The null hypothesis was rejected but Hypothesis 2e was also rejected for the CCTST as no correlation was predicted. The Pearson  $r$  coefficient calculated between number of years of health care employment and the CTOE total was  $-.023$ , which was not significant at the  $.05$  level. The null was accepted and Hypothesis 2e was rejected for CTOE as a positive correlation with the discipline specific critical thinking skills test was predicted.
- f. Hypothesis 2f stated that LVN status would not correlate with critical thinking scores on a general critical thinking skills test or a discipline specific critical thinking test for associate degree nursing students. As LVN status is a dichotomous variable and the formula for the point-biserial correlation coefficient is the same as the Pearson correlation coefficient, the Pearson  $r$  coefficient was calculated between LVN status and the CCTST total score. The Pearson  $r$  was  $.087$ , which was not significant at the  $.05$  level. The Pearson  $r$  coefficient calculated between LVN status and the CTOE total was  $-.017$ , which was not significant at the  $.05$  level. The null hypothesis was

- accepted and Hypothesis 2f was accepted for both CCTST and CTOE as no correlation was predicted.
- g. Hypothesis 2g stated that there would be no correlation between gender and critical thinking scores on a general critical thinking skills test and a discipline specific critical thinking test for associate degree nursing students. The Pearson  $r$  coefficient calculated between gender (another dichotomous variable) and the CCTST total was .053, which was not significant at the .05 level. The Pearson  $r$  coefficient calculated between gender and the CTOE total was .077, which was not significant at the .05 level. The null hypothesis was accepted and Hypothesis 2g was accepted for CCTST and CTOE as no correlation was predicted.
- h. Hypothesis 2h stated that there would be no correlation between ethnicity and critical thinking scores on a general critical thinking skills test and a discipline specific skills test. Ethnicity as measured on the demographic survey was a nominal variable with six levels. In order to be able to calculate the Pearson correlation coefficient, it was converted to a dichotomous variable of non-minority and minority. The Pearson  $r$  coefficient calculated between ethnicity and the CCTST total was -.228, which was significant at the .01 level. Since 1 = non-minority and 2 = minority, these results show a negative correlation with minority status. The null was rejected and Hypothesis 2g was rejected for CCTST as no correlation was predicted. The Pearson  $r$  coefficient calculated between ethnicity and the CTOE total was -.70, which was not significant at the .05 level. The null was accepted and Hypothesis 2g was accepted for CTOE as no correlation was predicted.

- i. Hypothesis 2i stated that there would be a statistically significant correlation between the variable set including age, completion of a previous degree and GPA and critical thinking scores on a general critical thinking skills test and a discipline specific critical thinking test. Two multiple linear regression analyses were performed to test this hypothesis. In the first regression analysis, the CCTST total score was the dependent variable. The three independent variables identified in the hypothesis were entered into the regression equation. Of the three entered independent variables, only GPA and degree status had statistically significant predictive value for total scores on the CCTST. Table 8 presents the results of the multiple regression analysis for age, degree status, and GPA as predictors of CCTST total scores. The F-value of 11.869 exceeded the critical F-value. The null hypothesis was rejected and hypothesis 2i was accepted. The R square was only .079, meaning that these three independent variables explained only 7.9% of the variance in CCTST scores.

In the second regression analysis, the CTOE total score was the dependent variable. The same three independent variables were entered into the regression equation. Of the three variables, only GPA had significantly significant predictive value for total CTOE scores. Table 9 shows the results of this regression model.

The F value of 5.542 exceeded the critical F-value. The null hypothesis was rejected and Hypothesis 2i was accepted. The R-square for the regression equation was .039, meaning that GPA explained only 3.9% of the variance in CTOE scores.

Table 8

Regression Analysis with CCTST Scores and Independent Variables

Model Summary					
Multiple R		.282			
R Square		.079			
Adjusted R Square		.073			
Standard Error		3.9385			
Analysis of Variance					
	Sum of Squares	df	Mean Square		
Regression	552.325	3	184.108		
Residual	6406.241	413	15.511		
Total	6958.566	416			
F = 11.869*					
*p<.01					
Variable	B	Standard Error	Beta	t	Sig.
(Constant)	6.879	1.651		4.167	.000
GPA	2.540	.493	.246	5.156	.000
Degree Stat.	.891	.316	.134	2.818	.005
Age	-2.763E-02	.025	-.054	-1.125	.261

## Additional Findings

## Subscale Correlations

The CCTST total scores positively and significantly correlated with each of its five subtests (see Table 10). The CCTST total scores also positively and significantly correlated with the CTOE evaluation ( $r = .209$ ) and explanation subscales ( $r = .195$ ). The CCTST total scores negatively and significantly correlated with the number of questions missed for the CTOE evaluation, inference, and explanation subscales. The CCTST total scores did not significantly correlate with the CTOE analysis or interpretation subscales.



Table 9

Regression Analysis with CTOE Scores and Independent Variables

Model Summary					
Multiple R	.199				
R Square	.039				
Adjusted R Square	.032				
Standard Error	16.6541				
Analysis of Variance					
.	Sum of Squares	df	Mean Square		
Regression	4611.286	3	1537.095		
Residual	112330.76	405	277.360		
Total	116942.04	408			
F = 5.542*					
*p<.01					
Variable	B	Standard Error	Beta	t	Sig.
(Constant)	57.403	6.995		8.207	.000
GPA	8.356	2.095	.197	3.990	.000
Degree Stat.	.931	1.342	.034	.694	.488
Age	-6.543E-02	.104	-.031	-.627	.531

The CTOE total scores positively and significantly correlated with each of its five subscales. Table 11 shows this relationship. The CTOE total scores also significantly and negatively correlated with the number of questions missed for each of the subcategories. The CTOE total scores positively and significantly correlated with each of the CCTST subscales. The very weak correlations were .157 for analysis, .155 for evaluation, .211 for inference, .167 for deduction, and .166 for induction. All were significant at the .01 level.

Table 10

Correlation Between CCTST Total and Subtest Scores

	CCTST total
Analysis	.625
Evaluation	.809
Inference	.704
Deduction	.807
Induction	.726

Table 11

Correlation Between CTOE Total and Subtest Scores

	CTOE Total	
	Raw Score	# Questions Missed
Interpretation	.434	-.360
Analysis	.501	-.500
Evaluation	.407	-.405
Inference	.628	-.606
Explanation	.449	-.449

## Partial Correlations

Several partial correlations were conducted controlling for various intervening variables. The correlation between CCTST and CTOE increased from .238 to .2418 (significant at .01 level) when the partial correlation controlled for LVN status. The correlation between CCTST and CTOE totals increased to .2405 when the partial correlation controlled for health care employment experience.

### Stepwise Multiple Regression

A stepwise multiple linear regression was performed using all the independent variables listed in hypotheses 2a-h. The stepwise multiple linear regression procedure applied with the dependent variable CCTST selected to retain GPA, health care employment, degree status, and first generation in college status. These independent variables had statistically significant predictive value for the total scores on the CCTST. Table 12 represents the results of the stepwise multiple regression analysis for these independent variables as predictors of CCTST total scores. The F value of 12.636 exceeded the critical F-value. The stepwise multiple linear regression procedure applied through SPSS failed to select gender, ethnicity, and LVN status for the regression equation. These independent variables had no statistically significant predictive value for CCTST total scores. The R-square of .110 showed that the four variables retained in the equation predicted only 11% of the variance in CCTST scores.

A similar stepwise multiple linear regression was performed for the dependent variable of CTOE total scores using the same seven independent variables. The regression model chose to retain only one variable, GPA, and failed to select the other six variables. With an R-square of .038, GPA predicted only 3.8% of the variance in the CTOE total scores. Table 13 shows the results of the stepwise multiple regression analysis for CTOE scores. The F-value of 15.979 exceeded the critical F-value.

Table 12

Summary of Stepwise Multiple Regression Analysis for CCTST

Model Summary					
Multiple R	.332				
R Square	.110				
Adjusted R Square	.102				
Standard Error	3.8904				
Analysis of Variance					
	Sum of Squares	df	Mean Square		
Regression	764.980	4	191.245		
Residual	6159.874	407	15.135		
Total	6924.854	411			
F = 12.636*					
*p<.01					
Variable	B	Standard Error	Beta	t	Sig.
(Constant)	5.973	1.691		3.533	.000
GPA	2.327	.489	.225	4.755	.000
HC Exp.	-.118	.040	-.140	-2.962	.003
Degree Stat.	.767	.314	.116	2.440	.015
1 <sup>st</sup> Gen. Status	.878	.390	.107	2.252	.025

Table 13

Summary of Stepwise Multiple Regression Analysis for CTOE

Model Summary					
Multiple R	.196				
R Square	.038				
Adjusted R Square	.036				
Standard Error	16.7070				
Analysis of Variance					
	Sum of Squares	df	Mean Square		
Regression	4460.265	1	4460.265		
Residual	112207.96	402	279.124		
Total	116668.23	403			
F = 15.979*					
*p<.01					
Variable	B	Standard Error	Beta	t	Sig.
(Constant)	56.554	6.608		8.559	.000
GPA	8.353	2.089	.196	3.997	.000

## CHAPTER V

### SUMMARY, CONCLUSIONS & RECOMMENDATIONS

#### Introduction

The purpose of this study was to determine whether there was any concurrent validity between two tools measuring critical thinking, one a discipline specific test, the Arnett Critical Thinking Outcome Evaluation (CTOE) tool, and the other a general critical thinking skills test, the California Critical Thinking Skills Test (CCTST). In addition, a secondary purpose of the study was to determine if any of several variables correlated with critical thinking scores of associate degree nursing students. Pearson correlation coefficients and regression analyses were used to test the hypotheses.

#### Summary of Findings

This study examined the correlation between the total scores and various subscale scores of two critical thinking tests, the CCTST and the Arnett CTOE tool for a convenience sample of associate degree nursing students from nine community colleges in Texas. Table 14 shows a summary of the results of each section of hypothesis one. Only one of the null hypotheses was accepted and three of the four research hypotheses were correctly predicted and accepted.

Table 14

Summary of Results of Hypotheses One

Variable 1	Variable 2	Pearson (r)	Null Hypothesis Result	Predicted Corr.	Actual Corr.	Research Hypothesis Result
CCTST Total Scores	CTOE Total Scores	.238**	Rejected	+	+	Accepted
CCTST Analysis	CTOE Analysis & Interpretation	.048	Accepted	+	0	Rejected
CCTST Evaluation	CTOE Evaluation & Explanation	.210**	Rejected	+	+	Accepted
CCTST Inference	CTOE Inference	.099*	Rejected	+	+	Accepted

\*p < .05; \*\*p < .01

Hypothesis two investigated the correlations between critical thinking scores on the two tests used and various independent variables. Tables 15 and 16 show a summary of the correlations between critical thinking scores and intervening variables along with the results of the hypotheses and the predicted and actual directions of any correlations. For both the CCTST and the CTOE and intervening variables, four of the eight hypotheses were predicted correctly and were accepted. In summary, GPA was positively correlated with CCTST scores, first generation in college status, health care employment experience and minority status were negatively correlated with CCTST scores, and the other variables, age, LVN status, gender and additional degree status had no correlation with CCTST scores. For the CTOE test, GPA had a positive correlation with CTOE

Table 15

Summary of Results of Hypotheses Two for CCTST Scores

Variable	Corr. With CCTST Scores	Sig. Level	Null Hypothesis Result	Predicted Corr.	Actual Corr.	Research Hypothesis Result
Age	.002	.972	Accepted	+	0	Rejected
Completion of previous degree	.078	.106	Accepted	+	0	Rejected
GPA	.254	.000	Rejected	+	+	Accepted
First Generation Status	.135	.005	Rejected	-	-	Accepted
Health Care Experience	-.183	.000	Rejected	0	-	Rejected
LVN status	.087	.071	Accepted	0	0	Accepted
Gender	.053	.274	Accepted	0	0	Accepted
Ethnicity	-.228	.000	Rejected	0	-	Rejected

scores and all other variables showed no correlation.

Additional findings showed that the subscales of the CCTST were positively and significantly correlated with the CCTST total scores with Pearson correlation coefficients ranging from a low of .625 to a high of .807. The subscales of the CTOE were also positively and significantly correlated with the CTOE total scores with lower Pearson correlation coefficients ranging from a low of .434 to a high of .628. The Watson Glaser Critical Thinking Appraisal subtests showed correlations between .56 and .79, in the middle of the results of the CCTST and CTOE subtest correlations. In addition, the correlation between the CCTST and CTOE total scores was slightly increased as measured by partial correlations when controlled for LVN status or health care employment experience.



Table 16

Summary of Results of Hypothesis Two for CTOE Scores

Variable	Corr. With CTOE Scores	Sig. Level	Null Hypothesis Result	Predicted Corr.	Actual Corr.	Research Hypothesis Result
Age	-.007	.884	Accepted	+	0	Rejected
Completion of previous degree	.011	.817	Accepted	+	0	Rejected
GPA	.192	.000	Rejected	+	+	Accepted
First Generation Status	.060	.216	Accepted	-	0	Rejected
Health Care Experience	-.023	.639	Accepted	+	0	Rejected
LVN status	-.017	.728	Accepted	0	0	Accepted
Gender	.077	.113	Accepted	0	0	Accepted
Ethnicity	-.70	.154	Accepted	0	0	Accepted

A stepwise multiple regression analysis with the dependent variable of CCTST scores selected to retain GPA, health care employment experience, additional degree status and first generation in college status and to reject gender, age, LVN status and ethnicity. These variables predicted only 11% of the variance in CCTST scores. A stepwise multiple regression analysis with the dependent variable of CTOE scores selected to retain only GPA and reject all six other variables. GPA predicted only 3.8% of the total variance in CTOE scores.

## Discussion of Findings

Although three of the four null hypotheses for hypothesis one were rejected and three of the four research hypotheses were accepted, the correlations, although significant

and positive, were weaker than expected. There are several possibilities that might explain these findings. First, there may be a fundamental difference between a general critical thinking skills test and a discipline specific critical thinking test. Critical thinking has been noted to be highly contextually sensitive (Jacobs, 1995). Cross and Steadman (1996) noted that different disciplines foster unique aspects of critical thinking. Kingten-Andrews (1991) suggested that different disciplines emphasize certain generic skills differently. In an examination of the relationship between conceptualization of critical thinking of a group of nurse educators and critical thinking experts, Gordon (1997) found a significant difference between their perceptions. If nursing experts were used to develop the discipline specific test, this may explain some of the differences. One study found a weak but positive relationship ( $r = .21$ ) between the CCTST scores and clinical judgment of BSN students using the Clinical Decision Making in Nursing Scale (Sorensen-Bowles, 1997). The discipline specific tool may be measuring clinical judgment more than critical thinking because of the different contexts.

In addition, although the two tests used in this study were based on the same definition of critical thinking, they may not have matched the definition used by the schools of nursing, the beliefs of the faculty working with the students, or the activities used to develop critical thinking skills in the curricula. When studies have measured the correlation between two critical thinking tests, validity has often been lower than expected. This has been explained because they had different scopes and were based on different theoretical constructs. The CCTST and CTOE are based on the same theoretical construct, but they have different scopes with one being a general test and the other a discipline specific test. With subject specific tests, there has been a concern raised that

interest in and familiarity with the subject matter influences one's critical thinking scores (Kingten- Andrews, 1991). Informal feedback from instructors or deans who administered the two tests in this study showed that participating students were more comfortable with the nursing specific test. There is a question as to whether the discipline specific test measures nursing knowledge as well as critical thinking and whether the two can be measured separately. The CTOE requires that the student use nursing content that every graduate should know.

A second reason the correlations may be weak is due to the different formats used by the two tests. The CCTST uses the multiple-choice format, which is useful for testing simple forms of critical thinking. The multiple-choice format is not sufficient as the only measure of critical thinking (Norris, 1988). The Arnett CTOE uses the constructed response format. These responses allow for more than one legitimate answer to a problem and are thought to offer more valid information on one's critical thinking abilities. The Wasatch Test of Critical Thinking, a structured response test only marginally correlated with multiple choice critical thinking tests at .36, but higher than the correlation in this study (Gunn, 1993). The correlation between the Arnett CTOE and the Arnett Computer Adaptive Test, which is a multiple-choice test used to predict success on the licensing exam, was only .18. This was not unexpected as the tools test different things and use different formats (Arnett, 1998b).

Another reason that the correlation was low may be due to the complexity of the concept of critical thinking. Frisby (1991) noted that validity coefficients might be low based on the complexity and many dimensions of critical thinking. The faculty at

Alverno College believe that standardized objective tests cannot adequately measure a concept such as critical thinking.

Two of the subtest groupings did correlate positively and significantly, although weakly, but the analysis subscale of the CCTST did not correlate with the analysis and interpretation subscales of the CTOE. Perhaps the analysis and interpretation subscales of the CTOE differ more due to the nursing contexts. Some of the subtests significantly correlated with subtests with which they were not predicted to correlate with. For example, the CTOE interpretation score correlated with the CCTST inference score but not the CCTST analysis score, and the CTOE evaluation and explanation scores correlated with the CCTST analysis and inference scores as well as the evaluation scores. Facione and Facione (1994) noted that factor analysis performed to differentiate different critical thinking skills into factors would likely fail because the skills have probably been used interactively to answer items. The critical thinking process is not easily compartmentalized into distinct skills that can be easily measured.

The scores of all of the subtests for the CCTST and four of the subtests for the CTOE were fairly consistent. The CTOE evaluation subtest scores were much lower than the other subtests. This finding was consistent with the results of the 1998 pilot study. Since the sample did not do as poorly on the CCTST evaluation subtest, whatever is inherently different in the evaluation subtest of the CTOE must be determined.

The results of hypothesis two add to current knowledge about CCTST and CTOE and ADN students. GPA was the only independent variable which correlated significantly and positively but weakly with the scores of both the CCTST and the CTOE. Of the studies reviewed for this research, three showed that GPA positively

correlated with CCTST scores, and one did not show a correlation. In addition, first generation at college status, health care experience and ethnicity negatively correlated with CCTST scores. Rebeschi and Thompson (1998) also noted a significant correlation between ethnicity and CCTST scores. No correlation was found with LVN status, gender, age, and additional degree status. In previous CCTST studies, one showed a positive correlation with age and two showed no correlation. No previous studies showed a correlation with gender. Previous studies have not measured correlation of CCTST scores with LVN status, additional degree status, or first generation at college status. GPA was the only variable that correlated with CTOE scores. No studies are available yet to compare these results to.

When the stepwise multiple regression analyses were performed, the retained variables explained only 11% of the variance in critical thinking scores on the CCTST. For the CTOE, the retained variable explained less than 4% of the variance in scores. Determining what variables contribute to the remaining variance in scores may be difficult as other variables are harder to measure. Terezini, Springer, Pascarella, and Nora (1995) found that student level of involvement in and out of class explained 52% of the variance in critical thinking scores on the CAAP. Student level of involvement, performance in nursing courses, performance in clinical, the level of discussion about critical thinking amongst the faculty, the teaching methods used to promote critical thinking, and the critical thinking outcome measures are other variables that may be included in further studies.

## Conclusions

The following conclusions are drawn from the results of this study:

1. There is limited concurrent validity between the CCTST, a general critical thinking skills test, and the Arnett CTOE, a discipline specific critical thinking test.
2. GPA is the only statistically valid predictor of critical thinking scores on a general critical thinking skills test and a discipline specific skills test for ADN students.
3. Neither the CCTST or the CTOE should be used alone as a single measure of critical thinking in ADN students.

## Implications for Practice

The following implications for practice are suggested by the results of this study:

1. Due to the low correlation between the two critical thinking tests, nursing programs may want to consider following the recommendations of Norris and Ennis (1989) and evaluate critical thinking in both discipline specific and everyday life contexts. In order to measure critical thinking for the NLNAC accreditation criteria, both the CCTST and the CTOE may be used as part of multiple measures of critical thinking since they appear to measure different aspects of critical thinking in nursing students. Data collected from multiple areas will help to develop a composite picture of critical thinking.
2. Major curriculum revisions should not be made based solely on the results of the critical thinking tests used in this study. Group means should be evaluated and compared to national norms for several years to get trended aggregate data before major decisions are made regarding changes.

3. Based on the CCTST mean score of the students in this sample being at the 30<sup>th</sup> percentile in relation to university senior nursing students who would be taking the same licensure examination at the end of the semester, associate degree nursing programs should continue to work on developing critical thinking skills throughout the program.
4. Nursing program faculty should investigate what the concept evaluation on the CTOE measures and work on improving those scores if they plan to continue to use the CTOE to measure the critical thinking of their students.

#### Recommendations for Future Study

The following are recommended for future studies based on the results of this study:

1. Another fifteen questions have been piloted and validated for use on the Arnett CTOE Tool. This study should be repeated with ADN and BSN students using the 30 item Arnett CTOE.
2. This study should be replicated with at least 500 students so that CCTST norms for associate degree nursing students can be developed.
3. The study should be replicated with a nationally representative sample of associate degree nursing students.
4. A similar study should be conducted evaluating the concurrent validity between two discipline specific critical thinking tests for both ADN and BSN students.
5. A similar study should be conducted comparing the Arnett CTOE tool and the California Critical Thinking Dispositions Inventory.

6. A similar study should be done evaluating the concurrent validity between the Arnett CTOE and the Watson Glaser Critical Thinking Appraisal.
7. A similar study should be conducted looking at other variables which may contribute to variance in critical thinking scores of ADN students such as performance in nursing courses, performance in clinical, and student level of involvement in and out of class.
8. A study should be conducted which compares critical thinking scores of ADN students on a standardized test with faculty perceptions of the student's critical thinking ability as shown in clinical situations.
9. Investigation should be conducted to determine why students at certain ADN programs perform better on the general critical thinking test and others perform better on the discipline specific critical thinking test or why some schools perform better than others. In this study, most schools ranked within one or two positions on each test when compared to each other but one school ranked first on the CCTST and 8<sup>th</sup> on the CTOE and another ranked 3<sup>rd</sup> on the CTOE and 7<sup>th</sup> on the CCTST. Certain aspects that should be addressed, include the admission requirements for each school, the critical thinking abilities of the faculty at each school, the emphasis on critical thinking strategies, different curriculum models and completion of certain core curriculum general education courses.
10. A longitudinal study should be conducted evaluating which critical thinking teaching strategies show significant improvements on which of the two critical thinking tests for associate degree nursing students over a specified period of time.



11. A longitudinal study should be conducted comparing ADN graduates in the last semester of their nursing program and graduates who have been out of school for at least six months.
12. Since no longitudinal studies have been reported using ADN students and CCTST scores, a study should evaluate the CCTST scores of ADN students at various points in the program such as before nursing courses are started, at the end of the first year, and at the end of the second year.
13. A similar study should be done evaluating the correlation between the Arnett CTOE and the Clinical Decision Making in Nursing Scale.

APPENDIX A  
DEMOGRAPHIC SURVEY

Code #: \_\_\_\_\_

Research Study on Critical Thinking Assessment in ADN Students  
Demographic Sheet

Please complete the following information:

1. Gender (Check one):    \_\_\_\_\_ Female    \_\_\_\_\_ Male
2. Age: \_\_\_\_\_
3. Ethnicity (Check one):    \_\_\_\_\_ African American  
   \_\_\_\_\_ Asian or Pacific Islander  
   \_\_\_\_\_ Caucasian  
   \_\_\_\_\_ Hispanic  
   \_\_\_\_\_ Native American  
   \_\_\_\_\_ Other \_\_\_\_\_
4. How many years experience do you have working (employment) in a health care setting? \_\_\_\_\_
5. Are you a LVN?    \_\_\_\_\_ Yes    \_\_\_\_\_ No  
  
      If yes, did you go through a LVN to ADN transition course or program?  
      \_\_\_\_\_ Yes    \_\_\_\_\_ No
6. Have you completed any previous educational degrees? \_\_\_\_\_ Yes    \_\_\_\_\_ No  
  
      If yes, which degree(s) do you have?    \_\_\_\_\_ Associate Degree  
   \_\_\_\_\_ Bachelor's Degree  
   \_\_\_\_\_ Master's Degree
7. Are you, your siblings, or your cousins the first generation of your family to go to and graduate from college?    \_\_\_\_\_ Yes    \_\_\_\_\_ No
8. What is your current cumulative GPA? \_\_\_\_\_  
      (If you don't remember it exactly, put your best estimate)

Thank you for your participation in this research study.

APPENDIX B  
COVER LETTER TO PARTICIPANTS

November 15, 1998

Dear Associate Degree Nursing Student,

I am a student at the University of North Texas. I would like to invite you to participate in a research study that I am conducting for my dissertation. Your nursing school has agreed to assist me with this study by allowing their students to participate. The purpose of this study is to see if there is any correlation between the California Critical Thinking Skills Test, a general critical thinking skills test, and the Arnett Critical Thinking Outcome Evaluation tool, a discipline specific critical thinking test for nursing. The results and feedback will help your school and other associate degree nursing programs decide whether to use either of these critical thinking assessment tools in their evaluation plan to meet National League for Nursing accreditation criteria. Benefits for yourself include receiving feedback on your level of critical thinking compared to national norms. In addition, this will be an opportunity for you to meet one of the proposed NLN competencies for associate degree nurses, that of participation in research. Because the majority of the current literature on critical thinking in nursing is on BSN students, you will also be helping to advance current knowledge about critical thinking in ADN students.

Your participation in this research study will involve completion of a one page demographic survey and taking the two critical thinking assessment tools, the California Critical Thinking Skills Test and the Arnett Critical Thinking Outcome Evaluation tool before the end of the spring semester. Your instructor/director will be keeping a tracking sheet that relates your name with an assigned code number. It is important that this code number be attached to the demographic survey and each of the tests. At the conclusion of the research, your instructor/director will destroy the tracking sheet. The data collected will be kept in strictest confidence and you will not be identified by name. There is no personal risk or discomfort directly involved with this research. Participation is voluntary. If you are willing to participate, please sign the attached consent form. You are free to withdraw your consent and discontinue participation in this study at any time. A decision to withdraw from the study will not affect your position in the nursing program. If you have any questions or problems that arise in connection with participation in this study, please contact me at (972) 932-4309.

Sincerely,

Helen Reid, RN, MSN  
University of North Texas Doctoral Student

THIS PROJECT HAS BEEN REVIEWED BY THE UNIVERSITY OF NORTH TEXAS COMMITTEE FOR THE PROTECTION OF HUMAN SUBJECTS (Phone: (940) 565-3940)

APPENDIX C  
CONSENT FORM

## CONSENT TO PARTICIPATE

I, \_\_\_\_\_ agree to participate in the study of the correlation between a general critical thinking skills test and a discipline specific critical thinking test for associate degree nursing students. As a participant I understand that I will be asked to complete a demographic survey and take the Arnett Critical Thinking Outcome Evaluation and the California Critical Thinking Skills Test before the end of the spring semester. I give permission for my instructor/director to release my scores on the Arnett Critical Thinking Outcome Evaluation tool and send the demographic survey and scantron for the California Critical Thinking Skills Test to the researcher.

I understand that a code number will identify the demographic survey and test results. I understand that my instructor/director will be keeping a tracking sheet that relates my name with my assigned code number. At the conclusion of the research, my instructor/director will destroy the tracking sheet. I understand that the data collected will be kept in strictest confidence by the researcher and that neither my school nor I as an individual will be identified by name. I agree that any information obtained from this research may be used in any way thought best for publication or education.

I understand there is no personal risk or discomfort directly involved with this research and that I am free to withdraw my consent and discontinue participation in this study at any time. A decision to withdraw from the study will not affect my position in the nursing program.

If I have any questions or problems that arise in connection with my participation in this study, I should contact the instructor/director that gave this consent form to me, or Helen Reid, the researcher, at (972) 932-4309.

\_\_\_\_\_  
Date

\_\_\_\_\_  
Signature of Participant

THIS PROJECT HAS BEEN REVIEWED BY THE UNIVERSITY OF NORTH TEXAS COMMITTEE FOR THE PROTECTION OF HUMAN SUBJECTS (Phone: (940) 565-3940)

## REFERENCES

- Adams, M. H., Whitlow, J. F., Stover, L. M., & Johnson, K. W. (1996). Critical thinking as an educational outcome. Nurse Educator, 21(3), 23-32.
- Alfaro-LeFevre, R. (1995). Critical thinking in nursing. Philadelphia: Saunders.
- Allegretti, C. L. & Frederick, J. N. (1995). A model for thinking critically about ethical issues. Teaching of Psychology, 22(1), 46-48.
- Anastasi, A. (1982). Psychological testing. NY: MacMillan Publishing.
- Angelo, T. A. (1995). Classroom assessment for critical thinking. Teaching of Psychology, 22(1), 6-8.
- Arnett, P. (1998a, September). Arnett Critical Thinking Outcome Evaluation. Paper presented at the meeting of the National League for Nursing, Chicago, IL.
- Arnett, P. (1998b). Arnett Critical Thinking Outcome Evaluation. Manuscript in preparation.
- Arnett Development Corporation. (n.d.). Critical Thinking Outcome Evaluation [Online]. Available: <http://arnettdevcorp.com/page4.html> [1998, March 7].
- Arter, J. A. (1987). How to select a test of higher order thinking skills. Portland, OR: The Northwest Regional Educational Laboratory. (ERIC Document Reproduction Service No. ED 290 760)
- Baker, C. B. (1996). Reflective learning: A teaching strategy for critical thinking. Journal of Nursing Education, 35(1), 19-22.
- Banta, T. W., Lund, J. P., Black, K. E., & Oblander, F. W. (1996). Assessment in practice. San Francisco: Jossey-Bass.
- Barnett, R. (1997). Higher education: A critical business. Bristol, PA: The Society for Research into Higher Education & Open University Press.
- Bauwens, E. E. & Gerhard, G. G. (1987). The use of the Watson-Glaser Critical Thinking Appraisal to predict success in a baccalaureate nursing program. Journal of Nursing Education, 26, 278-281.



Beeken, J. E., Dale, M. L., Enos, M. F., & Yarbrough, S. (1997). Teaching critical thinking skills to undergraduate nursing students. Nurse Educator, 22(3), 37-39.

Behrens, P. J. (1996). The Watson-Glaser Critical Thinking Appraisal and academic performance of diploma school students. Journal of Nursing Education, 35(1), 34-36.

Berger, A. (1985). Review of the Watson-Glaser Critical Thinking Appraisal. In Mental Measurements Yearbook (9<sup>th</sup> ed.), [CD-ROM]. Available: Silver Platter/AN 09139066

Berger, M. C. (1984). Clinical thinking ability and nursing students. Journal of Nursing Education, 23, 306-308.

Bingaman, N. D. (1993). A descriptive study of critical thinking abilities of nursing students and registered nurses [CD-ROM]. Abstract from: Silver Platter File: Dissertation Abstracts Item: 1352995

Board of Nurse Examiners for the State of Texas. (1999). 1998 summary of students enrollment and graduation data. Austin, TX: Author.

Bowers, B. & McCarthy, D. (1993). Developing analytic thinking skills early in undergraduate education. Journal of Nursing Education, 32, 107-137.

Brigham, C. F. (1989). Critical thinking skills in nursing students progressing through a nursing curriculum [CD-ROM]. Abstract from: Silver Platter File: Dissertation Abstracts Item: 8918496

Brookfield, S. D. (1987). Developing critical thinkers. San Francisco: Jossey-Bass.

Brooks, K. L. & Shepherd, J. M. (1990). The relationship between clinical decision-making skills in nursing and general critical thinking abilities of senior nursing students in four types of nursing programs. Journal of Nursing Education, 29, 391-398.

Brooks, K. L. & Shepherd, J. M. (1992). Professionalism versus general critical thinking abilities of senior nursing students in four types of nursing curricula. Journal of Professional Nursing, 8, 87-95.

Brown, H. N. & Sorrell, J. M. (1993). Use of clinical journals to enhance critical thinking. Nurse Educator, 18(5), 16-19.

Calderone, A. B. (1997). Critical thinking strategies in the classroom: Those that did and did not work. Nurse Educator, 22(3), 16, 20, 26.

California Academic Press. (1996a). California Critical Thinking Disposition Inventory [On-line]. Available: <http://www.calpress.com/dispinv.html>

California Academic Press. (1996b). California Critical Thinking Skills Test [On-line]. Available: <http://www.calpress.com/skill.html>

California Academic Press. (1996c). Frequently asked questions. [On-line]. Available: <http://www.calpress.com/faqs.html>

Carlson-Catalano, J. (1992). Empowering nurses for professional practice. Nursing Outlook, 40, 139-142.

Carpenter, C. B. & Doig, J. C. (1998). Assessing critical thinking across the curriculum. New Directions for Teaching and Learning, 34(2), 33-46.

Chubinski, S. (1996). Creative critical-thinking strategies. Nurse Educator, 21(6), 23-27.

Christen, R., Angermeyer, J., Davison, M. L., & Anderson, K. (n.d.). Performance assessments of critical thinking: The reflective judgment approach [On-line]. Available: <http://carei.coled.umn.edu/ResearchPractice/v2n1/performa.htm> [1998, March 7].

Clocklin, G. J. (1996). An exploration of the relationship between critical thinking skills and preferred learning style of first year students [CD-ROM]. Abstract from: Silver Platter File: Dissertation Abstracts Item: 9600940

Coffman, W. E. (1988). Measurement of thinking skills – A historical perspective. (ERIC Document Reproduction Service No. ED 299 322)

Colucciello, M. L. (1997). Critical thinking skills and dispositions of baccalaureate nursing students – A conceptual model for evaluation. Journal of Professional Nursing, 13, 236-245.

Cromwell, L. S. (1992). Assessing critical thinking. New Directions for Community Colleges, 20(1), 37-50.

Cronbach, L. J. & Meehl, P. E. (1967). Construct validity in psychological tests. In D. N. Jackson & S. Messick (Eds.), Problems in human assessment. NY: McGraw Hill.

Cross, K. P. & Steadman, M. H. (1996). Classroom research. San Francisco: Jossey Bass.

Davis, A. (1991). Higher order thinking skills compilation. Portland, OR: Northwest Regional Educational Library.

del Bueno, D. J. (1994). Why can't new grads think like nurses? Nurse Educator, 19(4), 9-11.

Dewey, J. (1933). How we think, a restatement of the relation of reflective thinking to the educative process. Boston, MA: Heath.

Dexter, P., Applegate, M., Backer, J., Claytor, K., Keffer, J., Norton, B., & Ross, B. (1997). A proposed framework for teaching and evaluating critical thinking in nursing. Journal of Professional Nursing, 13, 160-167.

Dobrzykowski, T. M. (1994). Teaching strategies to promote critical thinking skills in nursing staff. The Journal of Continuing Education in Nursing, 25, 272-276.

Dungan, J. M. (1985). Relationship of critical thinking and nursing process utilization (problem-solving) [CD-ROM]. Abstract from: Silver Platter File: Dissertation Abstracts Item: 8507848

Durgahee, T. (1996). Promoting reflection in post-graduate nursing: a theoretical model. Nurse Education Today, 16, 419-426.

Elliott, D. D. (1996). Promoting critical thinking in the classroom. Nurse Educator, 21(2), 49-51.

Ennis, R. H. (1985). A logical basis for measuring critical thinking skills. Educational Leadership, 43(2), 44-48.

Ennis, R. H. (1993). Critical thinking assessment. Theory into practice, 32(3), 179-186.

Facione, N. C. (1995). Critical thinking & clinical judgment: Goals 2000 for nursing science [On-line]. Available: <http://calpress.com/nurse2000.html>

Facione, N. C. (1997). Critical thinking assessment in nursing education programs: An aggregate data analysis. Millbrae, CA: California Academic Press.

Facione, N. C. & Facione, P. A. (1994). The California Critical Thinking Skills Test and the National League for Nursing accreditation requirement in critical thinking. (ERIC Document Reproduction Service No. ED 380 509)

Facione, N. C. & Facione, P. A. (1996a). Externalizing the critical thinking in knowledge development and clinical judgment. Nursing Outlook, 44, 129-136.

Facione, N. C. & Facione, P. A. (1996b). Holistic critical thinking scoring rubric [On-line]. Available: <http://www.calpress.com/rubric.html>

Facione, N. C., Facione, P. A., & Sanchez, C. A. (1994). Critical thinking disposition as a measure of competent clinical judgment: The development of the California Critical Thinking Disposition Inventory. Journal of Nursing Education, 33, 345-350.

Facione, P. A. (1990). Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. (ERIC Document Reproduction Service No. ED 315 428)

Facione, P. A. (1991). Using the California Critical Thinking Skills Test in research, evaluation, and assessment. (ERIC Document Reproduction Service No. ED 337 498)

Facione, P. A. (1996). Critical thinking: What it is and why it counts [On-line]. Available: <http://www.calpress.com/critical.html>

Facione, P. A. & Facione, N. C. (1998). The California Critical Thinking Skills Test Form A and B Test Manual. Millbrae, CA: California Academic Press.

Farley, M. J. & Elmore, P. B. (1992). The relationship of reading comprehension to critical thinking skills, cognitive ability, and vocabulary for a sample of underachieving college freshmen. Educational and Psychological Measurement, 52, 921-929.

Farrell, M. (1996). Planning for critical outcomes. Journal of Nursing Education, 35, 278-281.

Ford, J. S. & Profetto-McGrath, J. (1994). A model for critical thinking within the context of curriculum as praxis. Journal of Nursing Education, 33, 341-344.

Frederickson, K. & Mayer, G. G. (1977). Problem solving skills: What effect does education have? American Journal of Nursing, 77, 1167-1169.

Frisby, C. L. (1991). A meta-analytic investigation of the relationship between grade level and mean scores on the Cornell Critical Thinking Test (Level X). Measurement and Evaluation in Counseling and Development, 23, 162-170.

Frye, B., Alfred, N., & Campbell, M. (1999). Use of the Watson-Glaser Critical Thinking Appraisal. Nursing and Health Care Perspectives, 20(5), 253-255.

Geisinger, K. (1998). Review of the Watson-Glaser Critical Thinking Appraisal, Form S. In Mental Measurements Yearbook (13<sup>th</sup> ed.), [CD-ROM]. Available: Silver Platter/AN 13071819

Gendrop, S. C. & Eisenhauer, L. A. (1996). A transactional perspective on critical thinking. Scholarly Inquiry for Nursing Practice, 10, 329-343.

Glen, S. (1995). Developing critical thinking in higher education. Nurse Education Today, 15, 170-176.

Gokhale, A. A. (1995). Collaborative learning enhances critical thinking. Journal of Technology Education [On-line], 7(1), 1-7. Available <http://scholar.lib.vt.edu/ejournals/JTE/jte-v7n1/gokhale.jte-v7n1.html>

Gordon, J. M. (1997). Congruence between nursing educators and critical thinking experts on the conceptualization of critical thinking [CD-ROM]. Abstract from: Silver Platter File: Dissertation Abstracts Item: 9705235

Gross, Y. T., Takazawa, E. S., & Rose, C. L. (1987). Critical thinking and nursing education. Journal of Nursing Education, 26, 317-323.

Gunn, C. (1993). Assessing critical thinking: Development of a constructed-response test [CD-ROM]. Abstract from: Silver Platter File: Dissertation Abstracts Item: 93323932

Gunning, C. S. (1982). Relationships among field independence, critical thinking ability, and clinical problem-solving ability of baccalaureate nursing students [CD-ROM]. Abstract from: Silver Platter File: Dissertation Abstracts Item: 8128633

Hahnemann, B. K. (1986). Journal writing: A key to promoting critical thinking in nursing students. Journal of Nursing Education, 25, 213-215.

Haladyna, T. M. (1997). Writing test items to evaluate higher order thinking. Boston: Allyn and Bacon.

Halonen, J. S. (1995). Demystifying critical thinking. Teaching of Psychology, 22(1), 75-81.

Halpern, D. F. (1987). Student outcomes assessment: Introduction and overview. New Directions for Higher Education, 59(3), 5-8.

Halpern, D. F. (1989). Thought and knowledge: An introduction to critical thinking (2<sup>nd</sup> ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Halpern, D. F. (1993). Assessing the effectiveness of critical thinking instruction. The Journal of General Education, 42, 238-254.

Halpern, D. F. & Nummedal, S. G. (1995). Closing thoughts about helping students improve how they think. Teaching of Psychology, 22(1), 82-83.

Hanson, K. S. (1991). An analysis of the historical context of liberal education in nursing education from 1924 to 1939. Journal of Professional Nursing, 7, 341-350.

Hartley, D. & Aukamp, V. (1994). Critical thinking ability of nurse educators and nursing students. Journal of Nursing Education, 33(1), 34-35.

Helmstadter, G. (1985). Review of the Watson-Glaser Critical Thinking Appraisal. In Mental Measurements Yearbook (9<sup>th</sup> ed.), [CD-ROM]. Available: Silver Platter/AN 09139066

Heyman, G. A. & Daly, E. R. (1992). Teaching critical thinking in vocational-technical and occupational classes. New Directions for Community Colleges, 20(1), 103-108.

Hickman, J. S. (1993). A critical assessment of critical thinking in nursing education. Holistic Nurse Practitioner, 7(3), 36-47.

Hinkle, D., Wiersma, W. & Jurs, S. (1994). Applied statistics for the behavioral sciences (3<sup>rd</sup> ed.). Boston: Houghton Mifflin Company.

Holzemer, W. L. & McLaughlin, F. E. (1988). Concurrent validity of clinical simulations. Western Journal of Nursing Research, 10(1), 73-83.

Hopson, J. (1993). The differences in critical thinking ability between senior baccalaureate and associate degree nursing students [CD-ROM]. Abstract from: Silver Platter File: Dissertation Abstracts Item: 1355900

Howenstein, M. A., Bilodeau, K., Brogna, M. J. & Good, G. (1996). Factors associated with critical thinking among nurses. The Journal of Continuing Education in Nursing, 27, 100-103.

Hughes, J. (1995). Review of the Cornell Critical Thinking Tests. In Mental Measurements Yearbook (12<sup>th</sup> ed.), [CD-ROM]. Available: Silver Platter/AN 11070477

[InterEd@InterEd.com](http://www.interd.com/crit_thnk.htm) (n.d.). Critical thinking assessment [On-line]. Available: [http://www.interd.com/crit\\_thnk.htm](http://www.interd.com/crit_thnk.htm) [1998a, March 9].

[InterEd@InterEd.com](http://www.interd.com/ctn.htm) (n.d.). Critical Thinking Nursing – Version 2 [On-line]. Available: <http://www.interd.com/ctn.htm> [1998b, March 9].

[InterEd@InterEd.com](http://www.interd.com/ctn_2.htm) (n.d.). Critical Thinking Nursing – Version 2 [On-line]. Available: [http://www.interd.com/ctn\\_2.htm](http://www.interd.com/ctn_2.htm) [1998c, March 9].

Ircink Waite, R. M. (1989). A measurement of critical thinking in senior baccalaureate nursing students [CD-ROM]. Abstract from: Silver Platter File: Dissertation Abstracts Item: 9009974

Jacobs, P. M., Ott, B., Sullivan, B., Ulrich, Y., & Short, L. (1997). An approach to defining and operationalizing critical thinking. Journal of Nursing Education, 36(1), 19-22.

Jacobs, S. S. (1995). Technical characteristics and some correlates of the California Critical Thinking Skills Test, Forms A and B. Research in Higher Education, 36(1), 89-107.

James, C. R. & Clarke, B. A. (1994). Reflective practice in nursing: issues and implications for nurse education. Nurse Education Today, 14, 82-90.

Jones, S. A. & Brown, L. N. (1991). Critical thinking: impact on nursing education. Journal of Advanced Nursing, 16, 529-533.

Jones, S. A. & Brown, L. N. (1993). Alternative views on defining critical thinking through the nursing process. Holistic Nursing Practice, 7, 71-76.

Keeley, S., Browne, M., & Kreutzer, J. (1982). A comparison of freshmen and seniors on general and specific essay tests of critical thinking. Research in Higher Education, 17, 139-153.

Kelly, M. A. (1996). The effects of clinical experience and baccalaureate education in the critical thinking abilities of registered nurses [CD-ROM]. Abstract from: Silver Platter File: Dissertation Abstracts Item: 9611324

Kemp, V. H. (1985). Concept analysis as a strategy for promoting critical thinking. Journal of Nursing Education, 24, 382-384.

Kennison, M. & Brace, J. (1997). Digging deeper for creative solutions. Nursing 97, 97(9), 52-54.

Ketefian, S. (1981). Critical thinking, educational preparation, and development of moral judgment. Nursing Research, 30, 98-103.

King, P. M. & Kitchener, K. S. (1994). Developing reflective judgment. San Francisco: Jossey-Bass.

Kintgen-Andrews, J. (1988). Development of critical thinking: Career ladder P.N. and A.D. nursing students, pre-health science freshmen, and generic baccalaureate sophomore nursing students. (ERIC Document Reproduction Service No. ED 297 153)

Kintgen-Andrews, J. (1991). Critical thinking and nursing education: Perplexities and insights. Journal of Nursing Education, 30, 152-156.

Kokinda, M. A. (1990). The measurement of critical thinking skills in a selected baccalaureate nursing program [CD-ROM]. Abstract from: Silver Platter File: Dissertation Abstracts Item: 9004798

- Kramer, M. K. (1993). Concept clarification and critical thinking: Integrated processes. Journal of Nursing Education, 32, 406-414.
- Kuennen, J. & Moss, V. A. (1995). Community program planning: A clinical outcome. Journal of Nursing Education, 34, 387-389.
- Kumar, D. & Fritzer, P. (1994). Critical thinking assessment in a science and social studies context. American Secondary Education, 22(3), 2 – 4.
- Lacey, M. D. (1996). An examination of the disposition toward critical thinking and critical thinking ability of associate degree and baccalaureate nursing students and nursing faculty [CD-ROM]. Abstract from: Silver Platter File: Dissertation Abstracts Item: 9632064
- Lantz, J. M. (1986). Critical thinking through writing: Using personification to teach pharmacodynamics. Journal of Nursing Education, 25, 64-65.
- Lashley, M. & Wittstadt, R. (1993). Writing across the curriculum: An integrated curricular approach to developing critical thinking through writing. Journal of Nursing Education, 32, 422-424.
- Lenburg, C. B. & Mitchell, C. A. (1991). Assessment of outcomes: The design and use of real and simulation nursing performance examinations. Nursing & Health Care, 12, 68-72.
- Leppa, C. J. (1997). Standardized measures of critical thinking. Nurse Educator, 22(5), 29-33.
- Lewis, J. & Eakes, G. (1992). The AIDS care dilemma: An exercise in critical thinking. Journal of Nursing Education, 31, 136-137.
- Loevinger, J. (1967). Objectives tests as instruments of psychological theory. In D. N. Jackson & S. Messick (Eds.), Problems in human assessment. NY: McGraw Hill.
- Loving, G. L. (1993). Competence validation and cognitive flexibility: A theoretical model grounded in nursing education. Journal of Nursing Education, 32, 415-421.
- Lynch, M. H. (1989). Critical thinking: A comparative study of baccalaureate and associate degree nursing students [CD-ROM]. Abstract from: Silver Platter File: Dissertation Abstracts Item: 8820629
- McPeck, J. E. (1981). Critical thinking and education. New York: St. Martin's Press.
- Mackie, J. B. & Graham, P. B. (1996). Remodeling adult nursing. Nursing Forum, 31(4), 5-11.



Malcolm, K. (1995). Review of the Cornell Critical Thinking Tests. In Mental Measurements Yearbook (12<sup>th</sup> ed.), [CD-ROM]. Available: Silver Platter/AN11070477

Martin, G. W. (1996). An approach to the facilitation and assessment of critical thinking in nurse education. Nurse Education Today, 16, 3-9.

Matthews, C. A. & Gaul, A. L. (1979). Nursing diagnosis from the perspective of concept attainment and critical thinking. Advances in Nursing Science, 2(1), 17-26.

Maynard, C. A. (1996). Relationship of critical thinking ability to professional nursing competence. Journal of Nursing Education, 35(1), 12-18.

McMorris, R. (1995). Review of the California Critical Thinking Test. In Mental Measurements Yearbook (12<sup>th</sup> ed.), [CD-ROM]. Available: Silver Platter/AN 12071187

Michael, W. (1995). Review of the California Critical Thinking Test. In Mental Measurements Yearbook (12<sup>th</sup> ed.), [CD-ROM]. Available: Silver Platter/AN 12071187

Miller, M. A. (1992). Outcomes evaluation: measuring critical thinking. Journal of Advanced Nursing, 17, 1401-1407.

Miller, M. A. & Malcolm, N. S. (1990). Critical thinking in the nursing curriculum. Nursing & Health Care, 11, 67-73.

Modjeski, R. B. & Michael, W. B. (1983). An evaluation by a panel of psychologists of the reliability and validity of two tests of critical thinking. Educational and Psychological Measurement, 43, 1187-1197.

Morante, E. & Ulesky, A. (1984). Assessment of reasoning abilities. Educational Leadership, 42(1), 71-74.

Morrison, S., Smith, P., & Britt, R. (1996). Critical thinking and test item writing. Livingston, TX: Health Education Systems.

Moss, P. A. & Koziol, S. M. (1991). Investigating the validity of a locally developed critical thinking test. Educational Measurement: Issues and Practice, 10(3), 17-22.

Nathan, Y. H. (1997). Critical thinking: Impact on two classes of nursing students in an academic year [CD-ROM]. Abstract from: Silver Platter File: Dissertation Abstracts Item: 9734083

National Council of State Boards of Nursing. (1999). Age distribution by type of RN education program [On-line]. Available: [http://www.ncsbn.org/research/9/demographic/9/demo\\_rn\\_age.asp](http://www.ncsbn.org/research/9/demographic/9/demo_rn_age.asp)

National Council of State Boards of Nursing. (1999). RN gender and ethnicity [Online]. Available: [http://www.ncsbn.org/research/9/demographics/9/demo\\_rn\\_gender.asp](http://www.ncsbn.org/research/9/demographics/9/demo_rn_gender.asp)

National League for Nursing Accrediting Commission. (1997, Revised). Interpretive guidelines for standards and criteria. New York, NY: Author.

Neil, K. M., Lachat, M. F., & Taylor-Panek, S. (1997). Enhancing critical thinking with case studies and nursing process. Nurse Educator, 22(2), 30-32.

Norris, S. P. (1985). Synthesis of research on critical thinking. Educational Leadership, 42(8), 40-45.

Norris, S. P. (1987, October). Annotated list of critical thinking tests. Paper presented at the workshop on Assessing Higher Order Thinking Skills: Issues and Practices, Clackamas, OR.

Norris, S. P. (1988). Controlling for background beliefs when developing multiple-choice critical thinking tests. Educational Measurement: Issues and Practice, 7(3), 5-11.

Norris, S. P. (1990). Effect of eliciting verbal reports of thinking on critical thinking test performance. Journal of Educational Measurement, 27(1), 41-58.

Norris, S. P. & Ennis, R. H. (1989). Evaluating critical thinking. Pacific Grove, CA: Midwest Publications.

Nummedal, S. G. (1991). Designing a process to assess higher order thinking skills in college graduates: Issues of concern. (ERIC Document Reproduction Service No. ED 340 761)

Obenauf, P. (1996). A measurement of critical thinking dispositions and skills of baccalaureate nursing students [CD-ROM]. Abstract from: Silver Platter File: Dissertation Abstracts Item: 9543825

Oermann, M. H. (1997). Evaluating critical thinking in clinical practice. Nurse Educator, 22(5), 25-28.

Ory, J. C. (1991). Suggestions for deciding between commercially available and locally developed assessment instruments. NCA Quarterly, 66, 451-457.

O'Sullivan, P. S., Blevins-Stephens, W. L., Smith, F. M., & Vaughan-Wrobel, B. (1997). Addressing the National League for Nursing critical-thinking outcome. Nurse Educator, 22(1), 23-29.

Owen, S. (1998). Review of the Collegiate Assessment of Academic Proficiency. In Mental Measurements Yearbook (13<sup>th</sup> ed.), [CD-ROM]. Available: Silver Platter/AN 13011394

Pardue, S. F. (1987). Decision-making skills and critical thinking ability among associate degree, diploma, baccalaureate, and master's-prepared nurses. Journal of Nursing Education, 26, 354-360.

Paul, R. (1985). Critical thinking research: A response to Stephen Norris. Educational Leadership, 42(8), 46.

Paul, R. (1990). Critical thinking: What every person needs to survive in a rapidly changing world. Rohnert Park, CA: Center for Critical Thinking and Moral Critique.

Paul, R. (1992). Critical thinking: What, why, and how. New Directions for Community Colleges, 20(1), 3-23.

Paul, R. & Elder, L. (1996). The International Center for the Assessment of Higher Order Thinking [On-line]. Available: <http://www.sonoma.edu/cthink.icat.nclk>

Paul, R. & Elder, L. (n.d.) Three definitions of critical thinking [On-line]. Available: <http://education.lanl.gov/resources/cif/criticalthinkingdefinitions.htm> [1998, March 7].

Paul, R. & Nosich, G. (1992). A model for the national assessment of higher order thinking. (ERIC Document Reproduction Service No. ED 353 296)

Payne, D. E., Vowell, F. N., & Black, L. C. (1991). Assessment approaches in evaluation processes. NCA Quarterly, 66, 444-450.

Pearson, C. (1991). Barrier to success: Community college students critical thinking skills. (ERIC Document Reproduction Service No. ED 340 415)

Pepa, C. A., Brown, J. M., & Alverson, E. M. (1997). A comparison of critical thinking abilities between accelerated and traditional baccalaureate nursing students. Journal of Nursing Education, 36(1), 46-48.

Perciful, E. G. & Nester, P. A. (1996). The effect of an innovative clinical teaching method on nursing students' knowledge and critical thinking skills. Journal of Nursing Education, 35(1), 23-28.

Pless, B. S. & Clayton, G. M. (1993). Clarifying the concept of critical thinking in nursing. Journal of Nursing Education, 32, 425-428.

Polifroni, E. C. (1981). Problem-solving ability of senior nursing students in pre-service and inservice baccalaureate programs [CD-ROM]. Abstract from: Silver Platter File: Dissertation Abstracts Item: 8120847

Poteet, J. (1989). Review of the Ennis-Weir Critical Thinking Essay Test. In Mental Measurements Yearbook (10<sup>th</sup> ed.), [CD-ROM]. Available: Silver Platter/AN 10070141

Rane-Szostak, D. & Roberston, J. F. (1996). Issues in measuring critical thinking: Meeting the challenge. Journal of Nursing Education, 35(1), 5-11.

Richards, M. A. (1977). One integrated curriculum: An empirical evaluation. Nursing Research, 26, 90-95.

Russaw, E. H. (1997). The personally perceived problem technique: Enhancing clinical instruction. Nurse Educator, 22(4), 36-40.

Saarmann, L., Freitas, L., Rapps, J., & Riegel, B. (1992). The relationship of education to critical thinking ability and values among nurses: Socialization into professional nursing. Journal of Professional Nursing, 8(1), 26-34.

Saucier, B. L. (1995). Critical thinking skills of baccalaureate nursing students. Journal of Professional Nursing, 11, 351-357.

Schank, M. J. (1990). Wanted: Nurses with critical thinking skills. The Journal of Continuing Education in Nursing, 21(2), 86-89.

Schumacher, J. & Severson, A. (1996). Building bridges for future practice: An innovative approach to foster critical thinking. Journal of Nursing Education, 35(1), 31-33.

Scoloveno, M. A. (1981). Problem-solving ability of senior nursing students in three program types [CD-ROM]. Abstract from: Silver Platter File: Dissertation Abstracts Item: 8120858

Scriven, M. & Paul, R. (1996). Defining critical thinking [On-line]. Available: <http://www.sonoma.edu/cthink/University/univclass/Defining.nclk>

Sedlak, C. A. & Ludwick, R. (1996). Dressing up nursing diagnoses. Nurse Educator, 21(4), 19-22.

Sietsema, D. L. (1993). A comparison of critical thinking ability and clinical judgment skills in associate and baccalaureate senior nursing students [CD-ROM]. Abstract from: Silver Platter File: Dissertation Abstracts Item: 1351123

Smith, J. (1998). Review of the Collegiate Assessment of Academic Performance. In Mental Measurements Yearbook (13<sup>th</sup> ed.), [CD-ROM]. Available: Silver Platter/AN 13011394

Smith, M. L. (1995). A quantitative analysis of critical thinking abilities, learning and study strategies, and academic achievement in associate degree nursing students [CD-ROM]. Abstract from: Silver Platter File: Dissertation Abstracts Item: 9602082

Snyder, M. (1993). Critical thinking: A foundation for consumer-focused care. The Journal of Continuing Education in Nursing, 24, 206-210.

Sorensen Bowles, K. (1997). Effects of critical thinking, grade point average, and college experience on the clinical judgment skills of baccalaureate nursing students [CD-ROM]. Abstract from: Silver Platter File: Dissertation Abstracts Item: 9734861

Stadt, R. L. (1995). Critical thinking: A comparative study of baccalaureate, associate and certificate radiologic technology students [CD-ROM]. Abstract from: Silver Platter File: Dissertation Abstracts Item: 9513692

Sternberg, R. J. (1985). A statewide approach to measuring critical thinking skills. Educational Leadership, 43(2), 40-43.

Sternberg, R. J. (1986). Critical thinking: Its nature, measurement, and improvement. (ERIC Document Reproduction Service No. ED 272 882)

Sullivan, E. J. (1987). Critical thinking, creativity, clinical performance, and achievement in RN students. Nurse Educator, 12(2), 12-16.

Tanner, C. A. (1993). More thinking about critical thinking and clinical decision making. Journal of Nursing Education, 32, 387.

Tanner, C. A. (1996). Critical thinking revisited: Paradoxes and emerging perspectives. Journal of Nursing Education, 35(1), 3-4.

Tanner, C. A. (1997). Spock would have been a terrible nurse (and other issues related to critical thinking in nursing). Journal of Nursing Education, 36(1), 3-4.

Terenzini, P. T., Springer, L., Pascarella, E. T., & Nora, A. (1995). Influences affecting the development of students' critical thinking skills. Research in Higher Education, 36(1), 23-37.

Thompson, C. & Rebesch, L. (1998, September). Critical thinking skills and dispositions of baccalaureate nursing students at program entry and exit. Paper presented at the meeting of the National League for Nursing, Chicago, IL.

Thorndike, R. L. (1967). Reliability. In D. N. Jackson & S. Messick (Eds.), Problems in human assessment. NY: McGraw Hill.

Tompkins, G. (1989). Review of the Ennis-Weir Critical Thinking Essay Test. In Mental Measurements Yearbook (10<sup>th</sup> ed.), [CD-ROM]. Available: Silver Platter/AN 10070141

Tucker, R. W. (1996a). Less than critical thinking, Part II. Adult Assessment Forum [On-line], 6(4), 1-5. Available: [http://www.intered.com/v6n4\\_ed.htm](http://www.intered.com/v6n4_ed.htm)

Tucker, R. W. (1996b). The editor's desk – Less than critical thinking. Adult Assessment Forum [On-line], 6(3), 1-6. Available: [http://www.intered.com/v6n3\\_ed.htm](http://www.intered.com/v6n3_ed.htm)

Ulrich, D. & Glendon, K. (1995). Jigsaw: A critical-thinking experience. Nurse Educator, 20(3), 6-7.

Vanetzian, E. & Corrigan, B. (1996). "Prep" for class and class activity. Nurse Educator, 21(2), 45-48.

Vaughan-Wrobel, B. C., O'Sullivan, P., & Smith, L. (1997). Evaluating critical thinking skills of baccalaureate nursing students. Journal of Nursing Education, 36(10), 485-488.

Videbeck, S. L. (1997a). Critical thinking: A model. Journal of Nursing Education, 36(1), 23-28.

Videbeck, S. L. (1997b). Critical thinking: Prevailing practice in baccalaureate schools of nursing. Journal of Nursing Education, 36(1), 5-10.

Weddle, P. (n.d.). Do college seniors know how to think critically? [On-line]. Available: <http://www.csus.edu/ctl/ct.html> [1998, March 5].

White, N. E., Beardslee, N. Q., Peters, D., & Supples, J. M. (1990). Promoting critical thinking skills. Nurse Educator, 15(5), 16-19.

Whiteside, C. (1997). A model for teaching critical thinking in the clinical setting. Dimensions of Critical Care Nursing, 16, 152-161.

Whitney, D. R. (1993). Performance assessment: Lessons from history. Adult Learning, 4(3), 23-25.

Wissmann, J. E. (1996). Strategies for teaching critical thinking in pharmacology. Nurse Educator, 21(1), 42-46.

Woods, J. H. (1993). Affective learning: One door to critical thinking. Holistic Nursing Practice, 7(3), 64-70.

Yarbrough, D. B. (1992). Some lessons to be learned from a decade of general education outcomes assessment with the ACT COMP measures. Innovative Higher Education, 16, 223-231.

Zbilut, J. P. (1995). More on critical thinking, clinical judgment, and documentation. Journal of Emergency Nursing, 21, 483-484.

With Pearson correlation, three of four hypotheses concerning correlation between CCTST and CTOE scores were accepted, showing weak but significant correlation. GPA positively correlated but healthcare employment experience, first generation and minority status negatively correlated with CCTST scores. GPA correlated positively with CTOE scores. Stepwise multiple linear regression with CCTST scores retained GPA, healthcare employment experience, prior degree, and first generation in college status.

The significant, positive correlation between CCTST and CTOE scores was weaker than expected. This may be due to the different formats of the tools, or a fundamental difference between a general critical thinking skills test and a discipline specific tool. Critical thinking is highly contextually sensitive and disciplines emphasize skills differently. Both tests may be useful in a critical thinking assessment program since they measure different aspects and contribute to a composite picture of critical thinking. Research should continue on discipline specific tools.